DOCUMENT RESUME

ED 471 347 TM 034 638

AUTHOR	Onwuegbuzie, Anthony J.; Leech, Nancy L.
TITLE	Post-Hoc Power: A Concept Whose Time Has Come.
PUB DATE	2002-11-00
NOTE	40p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Chattanooga, TN, November 6-8, 2002).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	*Educational Research; *Power (Statistics); *Statistical Analysis
IDENTIFIERS	A Priori Tests; *Post Hoc Tests

ABSTRACT

This paper advocates the use of post-hoc power analyses. First, a history and definition of statistical power are provided. Second, reasons for the nonuse of a priori power analyses are presented. Post-hoc power is defined and its usefulness delineated. Fourth, a heuristic example is provided to illustrate how post-hoc power can help rule in/out rival explanations in the presence of statistically nonsignificant findings. (Contains 1 figure and 92 references.) (Author/SLD)

Running head: POST-HOC POWER

Post-Hoc Power: A: Concept Whose Time has Come

Anthony J. Onwuegbuzie

Howard University

Nancy L. Leech

University of Colorado at Denver

Correspondence should be addressed to Anthony J. Onwuegbuzie, Department of Human Development and Psychoeducational Studies, School of Education, Howard University, 2441 Fourth Street, NW, Washington, DC 20059, or E-Mail: (tonyonwuegbuzie@aol.com)

Paper presented at the annual meeting of the Mid-South Educational Research Association, Chattanooga, TN, November 6, 2002.

Abstract

The present paper advocates the use of *post-hoc* power analyses. First, a history and definition of statistical power are provided. Next, reasons for the non-use of *a priori* power analyses are presented. Third, *post-hoc* power is defined and its utility delineated. Finally, a heuristic example is provided to illustrate how *post-hoc* power can help to rule in/out rival explanations in the presence of statistically non-significant findings.

Post-Hoc Power: A Concept Whose Time has Come

For more than 75 years, null hypothesis significance testing (NHST) has dominated the quantitative paradigm, stemming from the seminal works of Fisher (1925/1941) and Neyman and Pearson (1928). NHST was designed to provide a means of ruling out a chance finding. Thereby reducing the chance of falsely rejecting the null hypothesis in favor of the alternative hypothesis (i.e., committing a Type I error). .

Although NHST has permeated the behavioral and social science field since its inception, its practice has been subjected to severe criticism, with the number of critics growing throughout the years. The most consistent criticism of NHST that has emerged is that statistical significance is not synonymous with practical significance. More specifically, statistical significance does not provide any information about how important or meaningful an observed finding is (e.g., Bakan, 1966; Cahan, 2000; Carver, 1978, 1993; Cohen, 1994, 1997; Guttman, 1985; Loftus, 1996; Meehl, 1967, 1978; Nix & Barnette, 1998; Onwuegbuzie & Daniel, in press; Rozeboom, 1960; Schmidt, 1992; 1996; Schmidt & Hunter, 1997)

As a result of this limitation of NHST, some researchers (e.g., Carver, 1993) contend that effect sizes, which represent measures of practical significance, should replace statistical significance testing completely. However, reporting and interpreting only effect sizes could lead to the over-interpretation of a finding. As noted by Robinson and Levin (1997):

> ...although effect sizes speak loads about the magnitude of a difference or relationship, they are, in and of themselves, silent with respect to the probability that the estimated difference or relationship is due to chance (sampling error). Permitting authors to promote and publish seemingly 'interesting' or 'unusual' outcomes when it can be documented that such outcomes are not really that unusual would open the publication floodgates to chance occurrences and other strange phenomena. (p. 25)

Onwuegbuzie (2001) calls the interpretation of a large effect size that represents a mere chance (i.e., statistically non-significant) finding a Type A error. In order to avoid such an error, Robinson and Levin (1997) proposed what they termed a "two-step process" for making statistical inferences. According to this model, a statistical significant observed finding is followed by the reporting and interpreting of one or more indices of practical significance; however, no effect sizes are reported in light of a statistically non-significant finding. In other words, analysts should determine first whether the observed result was statistically significant (Step 1), and, if and only if statistical significance is found, then they should report how large or important the observed finding is (Step 2). In this way, the statistical significance test in Step 1 serves as a gatekeeper for the reporting and interpreting of effect sizes in Step 2.

This two-step process is indirectly endorsed by the latest edition of the American Psychological Association (APA, 2001) Publication Manual:

> When reporting inferential statistics (e.g., *t* tests, *F* tests, and chi-square), *include* information about the obtained magnitude or value of the test

> statistic, the degrees of freedom, the probability of obtaining a value as extreme as or more extreme than the one obtained, and the direction of the effect. Be sure to *include* sufficient descriptive statistics (e.g., per-cell sample size, means, correlations, standard deviations) so that the nature of the effect being reported can be understood by the reader and for future meta-analyses. (p. 22)

Three pages later, APA (2001) states

> Neither of the two types of probability value directly reflects the magnitude of an effect or the strength of a relationship. For the reader 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)

On the following page, APA states that

> The general principle to be followed, however, is to provide the reader not only with information about statistical significance but also with enough information to assess the magnitude of the observed effect or relationship. (p. 26)

Most recently, Onwuegbuzie and Levin (2002) proposed a three-step procedure when two or more hypothesis tests are conducted within the same study, which involves testing the trend of the set of hypotheses at the third step. Using either the two-step method or the three-step method helps to reduce not only the probability of committing a Type A error, but also the probability of committing a Type B error, namely, declaring as important a statistically significant finding with a small effect size (Onwuegbuzie, 2001). However, whereas Type B error almost certainly will be reduced by using one of these methods compared to using NHST alone, the reduction in the probability of Type A error is not guaranteed using these procedures. This is because if statistical power is lacking, then the first step of the two-step method, and the first and third steps of the three-step procedure, which serve as "gatekeepers" for computing effect sizes, may lead to the non-reporting of a non-trivial effect (i.e., Type A error). Simply put, sample sizes that are too small increase the probability of a Type II error (not rejecting a false null hypothesis), and, subsequently, increase the probably of committing a Type B error.

Clearly, both error probabilities can be reduced if researchers conduct *a priori* power analyses in order to select appropriate sample sizes. However, unfortunately, such analyses are rarely employed (Cohen, 1992, Keselman, Huberty, Lix, Olejnik, Cribbie, Donahue, Kowalchuk, Lowman, Petoskey, Kesselman, & Levin, 1998; Onwuegbuzie, in press-a). When *a priori* power analyses have been omitted, researchers should conduct *post-hoc* power analyses, especially for non-statistically significant findings. This would help researchers determine whether low power threatens the internal validity of findings (i.e., Type A error). Yet, researchers typically have not used this technique.

Thus, this paper advocates the use of *post-hoc* power analyses. First, a history and definition of statistical power are provided. Next, reasons for the non-use of *a priori* power analyses are presented. Third, *post-hoc* power is defined and its utility delineated. Finally, a heuristic example is provided to illustrate how *post-hoc* power can help to rule in/out rival explanations in the presence of statistically non-significant findings.

History of Statistical Power

Up until the late 1920s, the statistical world was largely dominated by Sir Ronald A. Fisher, an eminent statistician and geneticist who developed an array of statistical techniques, most notably being the analysis of variance (Fisher, 1925/1941). Soon after Fisher's seminal work in 1925, J. Neyman and E.S. Pearson began to challenge some of Fisher's tenets. By the mid-1930s, a bitter debate emerged between the Fisherian school and the Neyman-Pearson school, which lasted until Fisher died in 1962 (Cowles, 1989). These debates largely pertained to issues of hypothesis testing in general and, in particular, the interpretation of statistical tests, the use of significance levels, and whether the declared level of statistical significance should be maintained throughout the research process (Chase & Tucker, 1976).

Fisher (Fisher, 1935, 1950, 1955) developed a comprehensive framework for drawing inferences from true experiments. Central to his framework was statistical tests. Fisher, who viewed statistical tests as significant tests (Chase & Tucker, 1976), developed the concept of the *null hypothesis*, which represented the assertion of no effect from the experimental treatments (although Fisher allowed for the testing of a specific non-zero value). Fisher (1935) posited that evidence against the null hypothesis would prevail when the observed experimental statistic (i.e., treatment difference) was so extreme compared to expected values that correspond to the hypothesized distribution for that statistic under the assumption that the null hypothesis is true that it was likely that the null hypothesis should be rejected. This was in essence the significance test.

Fisher (1935) observed that researchers deemed a finding under the null hypothesis to be "significant" when the result was more extreme than 95% of the values. Nevertheless, Fisher repeatedly noted that any decision to reject the null hypothesis was not irreversible. Also, he maintained that failure to reject the null hypothesis does not necessarily mean that the null hypothesis is true. That is, "the null hypothesis is never proved or established, but is possibly disproved, in the course of the experimentation. Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis" (Fisher, 1935, p. 19). Another important tenet of significance testing promoted by Fisher was that the significance test does not yield an actual probability for how true the hypothesis is--a misconception held by some researchers (Mulaik, Raju, & Harshman, 1997). More specifically, the probabilities involved with tests of significance "do not generally lead to any probability statements about the real world, but to a rational and well-defined measure of reluctance to the acceptance of the hypotheses they test" (Fisher, 1959, p. 44).

The Neyman-Pearson school initially started as an extension of Fisher's framework. These theorists published a series of papers (Neyman & Pearson, 1928, 1933a, 1933b; Pearson, 1941), whose impact continues today. In these articles, Neyman and Pearson treated statistical tests as decision tests. More specifically, they contended that significance tests should lead to the *accepting* or *rejecting* of the underlying hypothesis. These authors categorized hypotheses as being either simple or composite. A simple hypothesis represents the specification of a distinct point for a statistic among the set of all possible values that the statistic can take. On the other hand, a composite hypothesis denotes a range of values from the total sample space (Mulaik et al., 1997). According to Neyman and Pearson, the analyst's task is to divide the sample space into two regions, an acceptance region and a rejection region (i.e., critical region), and then make a decision as to whether to accept or to reject based on the region into which the observed value falls. The point that separates the acceptance and rejection region is called the critical value (cf. Kendall & Stewart, 1979). However, in order to determine the best critical region, the researcher must specify the probability of rejecting the hypothesis if the test statistic falls in the rejection region (i.e., is more extreme than is the critical value). This probability is the level of significance, or α. As advanced by Neyman and Pearson, the best critical region is that region of size α that also has the largest possible *power* of rejecting the null hypothesis assuming that the alternative hypothesis is true (Neyman & Pearson, 1928, 1933a, 1933b). As such the concept of power was born.

Definition of Statistical Power

Neyman and Pearson (1933b) were the first to discuss the concepts of Type I and Type II error. Type I error occurs when the researcher rejects the null hypothesis when it is true. As noted above, the Type I error probability is determined by the significance level (α). For example, if a 5% level of significance is designated, then the Type I error rate is 5%. Stated another way,

α represents the conditional probability of making a Type I error when the null hypothesis is true. Neymann and Pearson define α as the long-run relative frequency by which Type I errors are made over repeated samples from the same population under the same null and alternative hypothesis, assuming the null hypothesis is true. Conversely, Type II error occurs when the analyst accepts the null hypothesis when the alternative hypothesis is true. The conditional probability of making a Type II error under the alternative hypothesis is denoted by β.

Statistical power is the conditional probability of rejecting the null hypothesis (i.e., accepting the alternative hypothesis) when the alternative hypothesis is true. The most common definition of power comes from Cohen (1988), who defined the power of a statistical test as "the probability that it will lead to the rejection of the null hypothesis, i.e., the probability that it will result in the conclusion that the phenomenon exists" (p. 4). Power can be viewed as how likely it is that the researcher will find a relationship or difference that really prevails. It is given by 1 - β.

Statistical power estimates are affected by three factors. The first factor is level of significance. Holding all other aspects constant, increasing the level of significance increases power, but also increases the probability of rejecting the null hypothesis when it is true. The second influential factor is the effect size. Specifically, the larger the difference between the value of the parameter under the null hypothesis and the parameter under the alternative parameter, the greater the power to detect it. The third instrumental component is the sample size. The larger the sample size, the greater the likelihood of rejecting the null hypothesis (Chase & Tucker, 1976; Cohen, 1965, 1969, 1988, 1992).

Cohen (1965), in accordance with McNemar (1960), recommended a probability of .80 or greater for correctly rejecting the null hypothesis representing a medium effect at the 5% level of significance. This recommendation was based on considering the ratio of the probability of committing a Type I error (i.e., 5%) to the probability of committing a Type II error (i.e., 1 - .80 = .20). In this case, the ratio was 1:4, reflecting the contention that Type I errors are generally more serious than are Type II errors.

Power, level of significance, effect size, and sample size are related such that any one of these components is a function of the other three components. As noted by Cohen (1988), "when any three of them are fixed, the fourth is completely determined" (p. 14). Thus, there are four possible types of power analyses, in which one of the parameters is determined as a function of the other three, as follows: (a) power as a function of level of significance, effect size, and sample size; (b) effect size as a function of level of significance, sample size, and power; (c) level of significance as a function of sample size, effect size, and power; and (d) sample size as a function of level of significance, effect size, and power (Cohen, 1965, 1988). The latter type of power analysis is the most popular and most useful for planning research studies (Cohen, 1992). This form of power analysis, which is called an *a priori* power analysis, helps the researcher to ascertain the sample size necessary to obtain a desired level of power for a specified effect size and level of significance. Conventionally, most researchers set the power coefficient at .80 and the level of significance at .05. Thus, once the expected effect size and type of analysis are specified, then the sample size needed to meet all specifications can be determined.

The value of an *a priori* power analysis is that it helps the researcher in planning research studies (Sherron, 1988). By conducting such an analysis, researchers put themselves in the position to select a sample size that is large enough to lead to a rejection of the null hypothesis for a given effect size. Alternatively stated, *a priori* power analyses help researchers to obtain the necessary sample sizes to reach a decision with adequate power. Indeed, the optimum time to conduct a power analysis is during the research design phase (Wooley & Dawson, 1983).

Failing to consider statistical power can have dire consequences for researchers. First and foremost, low statistical power reduces the probability of rejecting the null hypothesis, and therefore, increases the probability of committing a Type II error (Bakan, 1966; Cohen, 1988), may increase the probability of committing a Type I error (Overall, 1969), may yield misleading results in power studies (Chase & Tucker, 1976), and may prevent potentially important studies from being published as a result of publication bias (Greenwald, 1975) and the "file-drawer problem," which represents the tendency to keep statistically non-significant results in file drawers (Rosenthal, 1979).

It has been exactly 40 years since Jacob Cohen (1962) conducted the first survey of power. In this seminal work, Cohen assessed the power of studies published in the abnormal-social psychology literature. Using the reported sample size and a non-directional significance level of 5%, Cohen calculated the average power to detect a hypothesized effect (i.e., hypothesized power) across the 70 selected studies for nine frequently used statistical tests, using small, medium, and large estimated effect-size values. The average power of the 2,088 major statistical tests were .18, .48, and .83 for detecting a small, medium, and large effect size, respectively. The average hypothesized statistical power of .48 for medium effects indicated that studies in the abnormal psychology field had, on average, less than a 50% chance of correctly rejecting the null hypothesis (Brewer, 1972; Halpin & Easterday, 1999).

During the next three decades after Cohen's (1962) investigation, several researchers have conducted hypothetical power surveys across a myriad of disciplines, including the following: applied and abnormal psychology (Chase & Chase, 1976), educational research (Brewer, 1972), educational measurement (Brewer & Owen, 1973), communication (Chase & Tucker, 1975; Katzer & Sodt, 1973), communication disorders (Kroll & Chase, 1975), mass communication (Chase & Baran, 1976), counselor education (Haase, 1974), social work education (Orme & Tolman, 1986), science education (Penick & Brewer, 1972; Woolley & Dawson, 1983), English education (Daly & Hexamer, 1983), gerontology (Levenson, 1980), marketing research (Sawyer & Ball, 1981), and mathematics education (Halpin & Easterday, 1999). The average hypothetical power of these 15 studies was .24, .63, and .85 for small, medium, and large effects, respectively. Assuming that a medium effect size is appropriate for use in most studies because of its combination of being practically meaningful and realistic (Cohen, 1965; Cooper & Findley, 1982; Haase, Waechter, & Solomon, 1982), the average power of .63 across these studies is disturbing. Similarly disturbing is the average hypothesized power of .64 for a medium effect reported by Rossi (1990) across 25 power surveys involving more than 1,500 journal articles and 40,000 statistical tests.

An even more alarming picture is painted by Schmidt and Hunter (1997), who reported that "the average [hypothesized] power of null hypothesis significance tests in typical studies and research literature is in the .40 to .60 range (Cohen, 1962, 1965, 1988, 1992; Schmidt, 1996; Schmidt, Hunter, & Urry, 1976; Sedlmeier & Gigerenzer, 1989)...[with] .50 as a rough average" (p. 40). Unfortunately, an average hypothetical power of .5, indicates that more than one-half of all statistical tests in the social and behavioral science literature will be statistically non-significant. As noted by Schmidt and Hunter (1997, p. 40), "This level of accuracy is so low that it could be achieved just by flipping a (unbiased) coin!" Yet, the fact that power is unacceptably low in most studies suggests that misuse of NHST is to blame, not the logic of NHST. Moreover, the publication bias that prevails in research suggests that the hypothetical power estimates provided above likely represent an upper bound. Thus, as declared by Rossi (1997), it is possible that "at least some controversies in the social and behavioral sciences may be artifactual in nature" (p. 178). Indeed, it can be argued that low statistical power represents more of a research design issue than it is a statistical issue, because it can be rectified by using a larger sample.

Bearing in mind the importance of conducting statistical power analyses, it is extremely surprising that very few researchers conduct and report power analyses for their studies (Brewer, 1972; Cohen, 1962, 1965, 1988, 1992; Keselman et al., 1998; Onwuegbuzie, in press-a; Sherron, 1988), even though statistical power has been promoted actively since the 1960s (Cohen, 1962, 1965, 1969), and even though for many types of statistical analyses (e.g., *r*, *z*, *F*, χ^2), tables have been provided by Cohen (1988, 1992) to determine the necessary sample size. Even when *a priori* power has been calculated, it is rarely reported (Woolley & Dawson, 1983). This lack of power analyses still prevails despite the recommendations of the APA (2001) to take power "seriously" and to "provide evidence that your study has sufficient power to detect effects of substantive interest" (p. 24).

The lack of use of power analysis might be the result of one or more of the following factors. First and foremost, evidence exists that statistical power is not sufficiently understood by researchers (Cohen, 1988, 1992). Second, it appears that the concept and applications of power are not taught in many undergraduate- and graduate-level statistical courses. Moreover, when power taught, it is likely that inadequate coverage is given. Disturbingly, Mundfrom, Shaw, Thomas, Young, and Moore (1998) reported that the issue of statistical power is regarded by instructors of research methodology, statistics, and measurement as being only the 34th most important topic in their fields out of the 39 topics presented. Also in this study, power received the same low ranking with respect to coverage in the instructors' classes. Clearly, if power is not being given a high status in quantitative-based research courses, then students similarly will not take it seriously. In any case, these students will not be suitably equipped to conduct such analyses.

Another reason for the spasmodic use of statistical power possibly stems from the incongruency between endorsement and practice. For instance, although APA (2001) stipulates that power analyses be conducted, despite providing several NHST examples, the manual does not provide any examples of how to report statistical power (Fidler, 2002). Harris (1997) also provides an additional rationale for the lack of power analyses:

> I suspect that this low rate of use of power analysis is largely due to the lack of proportionality between the effort required to learn and execute power analyses (e.g., dealing with noncentral distributions or learning the appropriate effect-size measure with which to enter the power tables in a given chapter of Cohen, 1977) and the low payoff from such an analysis (e.g., the high probability that resource constraints will force you to settle for a lower *N* than your power analysis says you should have)—especially given the uncertainties involved in a priori estimates of effect sizes and standard deviations, which render the resulting power calculation rather suspect. If calculation of the sample size needed for adequate power and for choosing between alternative interpretations of a nonsignificant result could be made more nearly equal in difficulty to the effort we've grown accustomed to putting into significance testing itself, more of us might in fact carry out these preliminary and supplementary analyses. (p. 165)

A further reason why *a priori* power analyses are not conducted likely stems from the fact that the most commonly used statistical packages, such as the Statistical Package for the Social Sciences (SPSS; SPSS Inc., 2001) and the Statistical Analysis System (SAS Institute Inc., 2002), do not allow researchers directly to conduct power analyses. Further, the statistical software programs that conduct power analyses (e.g., Erdfelder, Faul, & Buchner, 1996; Morse, 2001), although extremely useful, typically do not conduct other types of analyses, and thus researchers are forced to use at least two types of statistical software to conduct quantitative research studies, which is both inconvenient and possibly expensive. Even when researchers have power software in their possession, the lack of information regarding components needed to calculate power (e.g., effect size, variance) serves as an additional impediment to *a priori* power analyses.

It is likely that the lack of power analyses coupled with a publication bias promulgates the publishing of findings that are statistically significant but have small effect sizes (Type B error), as well as leading researchers to eliminate valuable hypotheses (Halpin & Easterday, 1999). Thus, we state in the strongest possible manner that all quantitative researchers conduct *a priori* power analyses whenever possible. These analyses should be reported in the Method section of research reports. This report also should include a rationale for criteria used for all input variables (i.e., power, significance level, effect size) (APA, 2001; Cohen, 1973, 1988). Inclusion of such analyses will help researchers to make optimum choices on the components (e.g., sample size, number of variables studied) needed to design a trustworthy study.

Post-Hoc Power Analyses

Whether or not an *a priori* power analysis is undertaken and reported, problems can still arise. One problem that commonly occurs in educational research is when the study is completed and a non-significant result is found. In many cases, the researcher then disregards the study (i.e., file-drawer problem) or when he/she submits the final report to a journal for review, finds it is rejected (i.e., publication bias). Unfortunately, most researchers do not determine whether the non-significant result is the result of insufficient statistical power. That is, without knowing the power of the statistical test, it is not possible to rule in or rule out low statistical power as a threat to internal validity (Onwuegbuzie, in press-b). Nor can an *a priori* power analysis necessarily rule in/out this threat. This is because *a priori* power analyses involve the use of *a priori* estimates of effect sizes and standard deviations (Harris, 1997). As such, *a priori* power analyses do not represent the power to detect the *observed* effect of the ensuing study; rather, they represent the power to detect *hypothesized* effects. Before the study is conducted, researchers do not know what the observed effect size will be. All they can do is try to estimate it based on previous research and theory (Wilkinson & the Task Force on Statistical Inference, 1999). The observed effect size could end up being much smaller or much larger than the hypothesized effect size on which the power analysis is undertaken. (Indeed, this is a criticism of the power surveys highlighted above; Mulaik et al., 1997.) In particular, if the observed effect size is smaller than what is proposed, the sample size yielded by the *a priori* power analysis might be smaller than is needed to detect it. In other words, a smaller effect size than anticipated increases the chances of Type II error.

On the other hand, the effect of power on a statistically non-significant finding can be assessed more appropriately by using the observed (true) effect to investigate the performance of a NHST (Mulaik et al., 1997; Schmidt, 1996; Sherron, 1988). Such a technique leads to what is often called a *post-hoc* power analysis. Interestingly, several authors have recommended the use of post-hoc power analyses for statistically non-significant findings (Cohen, 1969; Dayton, Schafer, & Rogers, 1973; Fagely, 1985; Fagley & McKinney, 1983; Sawyer & Ball, 1981; Woolley & Dawson, 1983).

When post-hoc power should be reported has been the subject of debate. While some researchers advocate that post-hoc power always be reported (e.g., Woolley & Dawson, 1983), the majority of researchers advocate reporting post-hoc power only for statistically non-significance results (Cohen, 1965; Fagely, 1985; Fagley & McKinney, 1983; Sawyer & Ball, 1981). However, both sets of analysts agree that estimating the power of significance tests that yield statistically non-significant findings plays an important role in their interpretation (e.g., Fagely, 1985; Fagley & McKinney, 1983; Sawyer & Ball, 1981; Tversky & Kahneman, 1971). Specifically, statistically non-significant results in a study with low power suggest ambiguity. Conversely, statistically non-significant results in a study with high power contribute to the body of knowledge because power can be ruled out as a threat to internal validity (e.g., Fagely, 1985; Fagley & McKinney, 1983; Sawyer & Ball, 1981; Tversky & Kahneman, 1971). To this end, statistically non-significant results can make a greater contribution to the research community than they presently do. As noted by Fagely (1985), "Just as rejecting the null does not guarantee large and meaningful effects, accepting the null does not preclude interpretable results" (p. 392).

Conveniently, post-hoc power analyses can be conducted relatively easily because some of the major statistical software programs compute post-hoc power estimates. In fact, post-hoc power coefficients are available in SPSS for the General Linear Model. For example, the post-hoc power procedure for analyses of variance and multiple analyses of variance is contained within the "options" button.

Framework for Conducting Post-Hoc Power Analyses

We agree that post-hoc power analyses should accompany statistically non-significant findings.[1] In fact, such analyses can provide useful information for replication studies. In particular, the components of the post-hoc power analysis can be used to conduct *a priori* power analyses in subsequent replication investigations.

Figure 1 displays our power-based framework for conducting NHST. Specifically, once the research purpose and hypotheses have been determined, the next step is to use an *a priori* power analysis to design the study. Once data have been collected, the next step is to test the hypotheses. For each hypothesis, if statistical significance is reached (e.g., at the 5% level), then the researcher should report the effect size and confidence interval around the effect size (e.g., Bird, 2002; Chandler, 1957; Cumming & Finch, 2001; Fleishman, 1980; Steiger & Fouladi, 1992, 1997; Thompson, 2002). Conversely, if statistical significance is not reached, then the researcher should conduct a post-hoc power analysis in an attempt to rule in or to rule out inadequate power (e.g., power $<$ .80) as a threat to the internal validity of the finding.

Insert Figure 1 about here

Heuristic Example

Recently, Onwuegbuzie, Witcher, Filer, Collins, and Downing (in press) conducted a study investigating characteristics associated with teachers' views on discipline. The theoretical framework for this investigation, though not presented here, can be found by examining the original study. Although several independent variables were examined by Onwuegbuzie et al., we will restrict our attention to one of them, namely, ethnicity (i.e., Caucasian-American vs. minority) and its relationship to discipline styles.

Participants were 201 students at a large mid-southern university who were either preservice (77.0%) or inservice (23.0%) teachers. The sample size was selected via an *a priori* power analysis because it provided acceptable statistical power (i.e., .82) for detecting a moderate difference in means (i.e., Cohen's [1988] d = .5) at the (two-tailed) .05 level of significance, maintaining a familywise error of 5% (i.e., approximately .01 for each set of statistical tests comprising the three subscales used) (Erdfelder et al., 1996). The preservice teachers were selected from several sections of an introductory-level undergraduate education class. On the other hand, the inservice teachers represented graduate students who were enrolled in one of two sections of a research methodology course.

On the first week of class, participants were administered the Beliefs on Discipline Inventory (BODI), which was developed by Roy T. Tamashiro and Carl D. Glickman (as cited in Wolfgang & Glickman, 1986). This measure was constructed to assess teachers' beliefs on classroom discipline by indicating the degree to which they are non-interventionists, interventionists, and interactionalists. The BODI contains 12 multiple-choice items, each with two response options. For each item, participants are asked to select the statement with which they most agree. The BODI contains three subscales representing the non-interventionist, interventionist, and interactionalist orientations, with scores on each subscale ranging from zero to eight. A high score on any of these scales represents a teacher's proclivity toward the particular discipline approach. For the present study, the non-interventionist, interventionist, and interactionalist subscales generated scores that had a classical theory alpha reliability coefficient of .72 (95% confidence interval [CI] = .66, .77), .75 (95% CI = .69, .80), and .94 (95% CI = .93, .95), respectively.

A series of independent *t*-tests, using the Bonferroni adjustment to maintain a familywise error of 5%, revealed no statistically significant difference between Caucasian-American and minority participants for scores on the Interventionist ($t = -1.47$, $p > .05$), Non-interventionist ($t = 0.88$, $p > .05$), and Interactionalist ($t = 0.52$, $p > .05$) subscales. After finding statistical non-significance, the researchers could have concluded that there were no ethnic differences in discipline beliefs. However, they decided to conduct a post-hoc power analysis. The post-hoc power analysis for this test of ethnic differences revealed low statistical power. Thus, these researchers concluded the following:

> The finding of no ethnic differences in discipline beliefs also is not congruent with Witcher et al. (2001), who reported that minority preservice teachers less often endorsed classroom and behavior management skills as characteristic of effective teachers than did Caucasian-American preservice teachers. Again, the non-significance could have stemmed from the relatively small proportion of minority students (i.e., 12.9%), which induced relatively low statistical power (i.e., 0.66) for comparing the two groups (Erdfelder et al., 1996). Replications are thus needed to determine the reliability of the present findings of no ethnic differences in discipline belief. (p. 19)

Thus, the post-hoc power analysis allowed the statistically non-significant finding pertaining to ethnicity to be placed in a more appropriate context.

Summary and Conclusions

Robinson and Levin (1997) proposed a two-step procedure for analyzing empirical data, whereby researchers first evaluate the probability of an observed effect (i.e., statistical significance) and, if and only if statistical significance is found, then they assess the effect size. Recently, Onwuegbuzie and Levin (2002) proposed a three-step procedure when two or more hypothesis tests are conducted within the same study, which involves testing the trend of the set of hypotheses at the third step. Although both methods are appealing, their effectiveness depend on the statistical power of the hypothesis tests. Specifically, if power is lacking, then the first step of the two-step method, and the first and third steps of the three-step procedure, which serve as "gatekeepers" for computing effect sizes, may lead to the non-reporting of a non-trivial effect (i.e., Type A error; Onwuegbuzie, 2001).

Because the typical level of power for medium effect sizes in the behavioral and social sciences is around .50 (Cohen, 1962), the incidence of Type A error likely is high. Clearly, this incidence can be reduced if researchers conduct an *a priori* power analysis in order to select appropriate sample sizes. However, such analyses are rarely employed (Cohen, 1992). Regardless, when a statistically non-significant finding emerges, researchers should then conduct a *post-hoc* power analysis. This would help researchers determine whether low power threatens the internal validity of their findings (i.e., Type A error). Yet, virtually no researcher has formally used this technique.

Thus, this paper advocates the use of *post-hoc* power analyses for statistically non-significant findings. First, a history and definition of statistical power were provided. Next, reasons for the non-use of *a priori* power analyses were presented. Third, *post-hoc* power was defined and its utility delineated. Finally, a heuristic example was provided to illustrate how *post-hoc* power can help to rule in/out rival explanations to observed findings.

Although we advocate the use of post-hoc power analyses in the presence of statistically non-significant results, we believe that such analyses should never be used as a substitute for *a priori* power analyses. Moreover, we recommend that *a priori* power analyses always be conducted and reported. Nevertheless, even when an *a priori* power analysis has been conducted, we believe that a post-hoc analysis also should be performed if one or more statistically non-significant findings emerge. Post-hoc power analyses rely more on available data and less on speculation than do *a priori* power analyses that are based on hypothesized effect sizes.

Indeed, we agree with Woolley and Dawson (1983), who suggest "editorial policies to require all such information relating to *a priori* design considerations and *post h*oc interpretation to be incorporated as a standard component of any research report submitted for publication" (p. 680). Although it could be argued that this recommendation is bold, it is no more bold than the editorial policies at 20 journals that now formally stipulate that effect sizes be reported for all statistically significant findings (Capraro & Capraro, 2002). In fact, post-hoc power provides a nice balance in report writing because we believe that post-hoc power is to statistically non-significant findings as effect sizes are to statistically significant findings. In any case, we believe that such a policy of conducting and reporting *a priori* and post-hoc power analyses would simultaneously reduce the incidence of Type II and Type A errors and, subsequently, reduce the incidence of publication bias and the file-drawer problem. This can only help to increase the accumulation of knowledge across studies because meta-analysts will have much more information to use. This surely would represent a step in the right direction.

References

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin, 6*, 432-437.

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

Brewer, J. K. (1972). On the power of statistical tests in the American Education Research Journal. *American Education 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.

Cahan, S. (2000). Statistical significance is not a "Kosher Certificate" for observed effects: A critical analysis of the two-step approach to the evaluation of empirical results. *Educational Researcher, 29*(1), 31-34.

Capraro, R. M., & Capraro, M. M. (2002). Treatments of effect sizes and statistical significance tests in textbooks. *Educational and Psychological Measurement, 62*, 771-782.

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

Chandler, R. E. (1957). The statistical concepts of confidence and significance.

Psychological Bulletin, 54, 429-430.

Chase, L. J., & Baran, S. J. (1976). An assessment of quantitative research in mass communication. *Journalism Quarterly, 53*, 308-311.

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-analytical 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. *The Psychological Record, 26*, 473-486.

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

Cohen, J. (1965). Some statistical issues in psychological research. In B. B. Wolman (Ed.), *Handbook of clinical psychology* (pp. 95-121). New York: McGraw-Hill.

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

Cohen, J. (1973). Statistical power analysis and research results. *American Educational Research Journal, 10*, 225-230.

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

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

Cohen, J. (1997). The earth is round ($p < .05$). In L.L. Harlow, S.A. Mulaik, & J.H. Steiger (Eds.), *What if there were no significance tests?* (pp. 21-35). Mahwah, NJ: Erlbaum.

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

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

Cumming, G., & Finch, S. (2001). A primer on the understanding, use and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement, 61*, 532-575.

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

Dayton, C. M., Schafer, W. D., & Rogers, B. G. (1973). On appropriate uses and interpretations of power analysis: A comment. *American Educational Research Journal, 10*, 231-234.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers, 28*, 1-11.

Fagely, N. S. (1985). Applied statistical power analysis and the interpretation of nonsignificant results by research consumers. *Journal of Counseling Psychology, 32*, 391-396.

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

Fidler, F. (2002). The fifth edition of the *APA Publication Manual:* Why its statistics recommendations are so controversial. *Educational and Psychological Measurement, 62*, 749-770.

Fisher, R. A. (1925/1941). *Statistical methods for research workers* (84th ed.) Edinburgh, Scotland: Oliver & Boyd. (Original work published in 1925).

Fisher, R. A. (1935). *The design of experiments*. London, Oliver and Boyd.

Fisher, R. A. (1950). *Contributions to mathematical statistics*. London: Chapman and Hall.

Fisher, R. A. (1955). Statistical methods and scientific induction. *Journal of the Royal Statistical Society, Series B, 17*, 69-78.

Fisher, R. A. (1959). *Statistical methods and scientific inference*. Edinburgh: Oliver & Boyd.

Fleishman, A. I. (1980). Confidence intervals for correlation ratios. *Educational and Psychological Measurement, 40*, 659-670.

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82*, 1-20.

Guttman, L. B. (1985). The illogic of statistical inference for cumulative science. *Applied Stochastic Models and Data Analysis, 1*, 3-10.

Haase, R. F. (1974). Power analysis of research in counselor education. *Counselor Education and Supervision, 14*, 124-132.

Haase, R. F., Waechter, D. M., & Solomon, G. S. (1982). How significant is a significant difference? Average effect size of research in counseling psychology. *Journal of Counseling Psychology, 29*, 58-65.

Halpin, R., & Easterday, K. E. (1999). The importance of statistical power for quantitative research: A post-hoc review. *Louisiana Educational Research Journal, 24*, 5-28.

Harris, R. J. (1997). Reforming significance testing via three-valued logic. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 145-174). Mahwah, NJ: Erlbaum.

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

Kendall, M., & Stuart, A. (1979). *The advanced theory of statistics: Inference and relationship* (Vol. 2). London: Charles Griffin & Co.

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, 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 analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research, 68*, 350-386.

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

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

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

McNemar, Q. (1960). At random: Sense and nonsense. *American Psychologist, 15*, 295-300.

Meehl, P. E. (1967). Theory testing in psychology and physics: A methodological paradox. *Philosophy of Science, 34*, 103-115.

Meehl, P. E. (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.

Morse, D. T. (2001, November). *YAPP: Yet another power program*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Chattanooga, TN

Mulaik, S. A., Raju, N. S., & Harshman, R. A. (1997). There is a time and a place for significance testing. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 65-115). Mahwah, NJ: Erlbaum.

Mundfrom, D. J., Shaw, D. G., Thomas, A., Young, S., & Moore, A. D. (1998, April). *Introductory graduate research courses: An examination of the knowledge base*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Neyman, J., & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for the purposes of statistical inference. *Biometrika, 20A*, 175-200, 263-294.

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

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

Nix, T. W., & Barnette, J. J. (1998). The data analysis dilemma: Ban or abandon. A review of null hypothesis significance testing. *Research in the Schools, 5*, 3-14.

Onwuegbuzie, A. J. (2001, November). *A new proposed binomial test of result direction.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Little Rock, AR.

Onwuegbuzie, A. J. (in press-a). Common analytical and interpretational errors in educational research: An analysis of the 1998 volume of the British Journal of Educational Psychology. *Research for Educational Reform.*

Onwuegbuzie, A. J. (in press-b). Expanding the framework of internal and external validity in quantitative research. *Research in the Schools.*

Onwuegbuzie, A. J., & Daniel, L. G. (in press). Typology of analytical and interpretational errors in quantitative and qualitative educational research. *Current Issues in Education.*

Onwuegbuzie, A. J., & Levin, J. R. (2002). *A new proposed three-step method for testing result direction.* Manuscript in preparation.

Onwuegbuzie, A. J., Witcher, A. E., Filer, J., Collins, K. M. T., & Downing, J. (in press). Factors associated with teachers' beliefs about discipline in the context of practice. *Research in the Schools, 10*(2).

Orme, J. G., & Tolman, R. M. (1986). The statistical power of a decade of social work education research. *Social Science Review, 60*, 619-632.

Overall, J. E. (1969). Classical statistical hypotheses within the context of Bayesian theory. *Psychological Bulletin, 71*, 285-292.

Pearson, E. S. (1941). A note on further properties of statistical tests. *Biometrika, 32*, 59-61.

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

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

Rosenthal, R. (1979). The "file-drawer problem" and tolerance for null results. *Psychological Bulletin, 86*, 638-641.

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.

Rossi, J. S. (1997). A case study in the failure of psychology as a cumulative science: The spontaneous recovery of verbal learning. In L. L. Harlow, S.

A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 175-197). Mahwah, NJ: Erlbaum.

Rozeboom, W. W. (1960). The fallacy of the null hypothesis significance test. *Psychological Bulletin, 57,* 416-428.

SAS Institute Inc. (2002). *SAS/STAT User's Guide* (Version 8.2) [Computer software]. Cary, NC: SAS Institute Inc.

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

Schmidt, F. L. (1992). What do data really mean? *American Psychologist, 47,* 1173-1181.

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

Schmidt, F. L., & Hunter, J. E. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. In L.L. Harlow, S.A. Mulaik, & J.H. Steiger (Eds.), *What if there were no significance tests?* (pp. 37-64). Mahwah, NJ: Erlbaum.

Schmidt, F. L., Hunter, J. E., & Urry, V. E. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology, 61*, 473-485.

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.

SPSS Inc. (2001). *SPSS 11.0 for Windows.* [Computer software]. Chicago, IL: SPSS Inc.

Steiger, J. H., & Fouladi, R. T. (1992). R^2: A computer program for interval estimation, power calculation, and hypothesis testing for the squared multiple correlation. *Behavior Research Methods, Instruments, and Computers, 4*, 581-582.

Steiger, J. H., & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L.L. Harlow, S.A. Mulaik, & J.H. Steiger (Eds.), *What if there were no significance tests?* (pp. 221-257). Mahwah, NJ: Erlbaum.

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

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105-110.

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

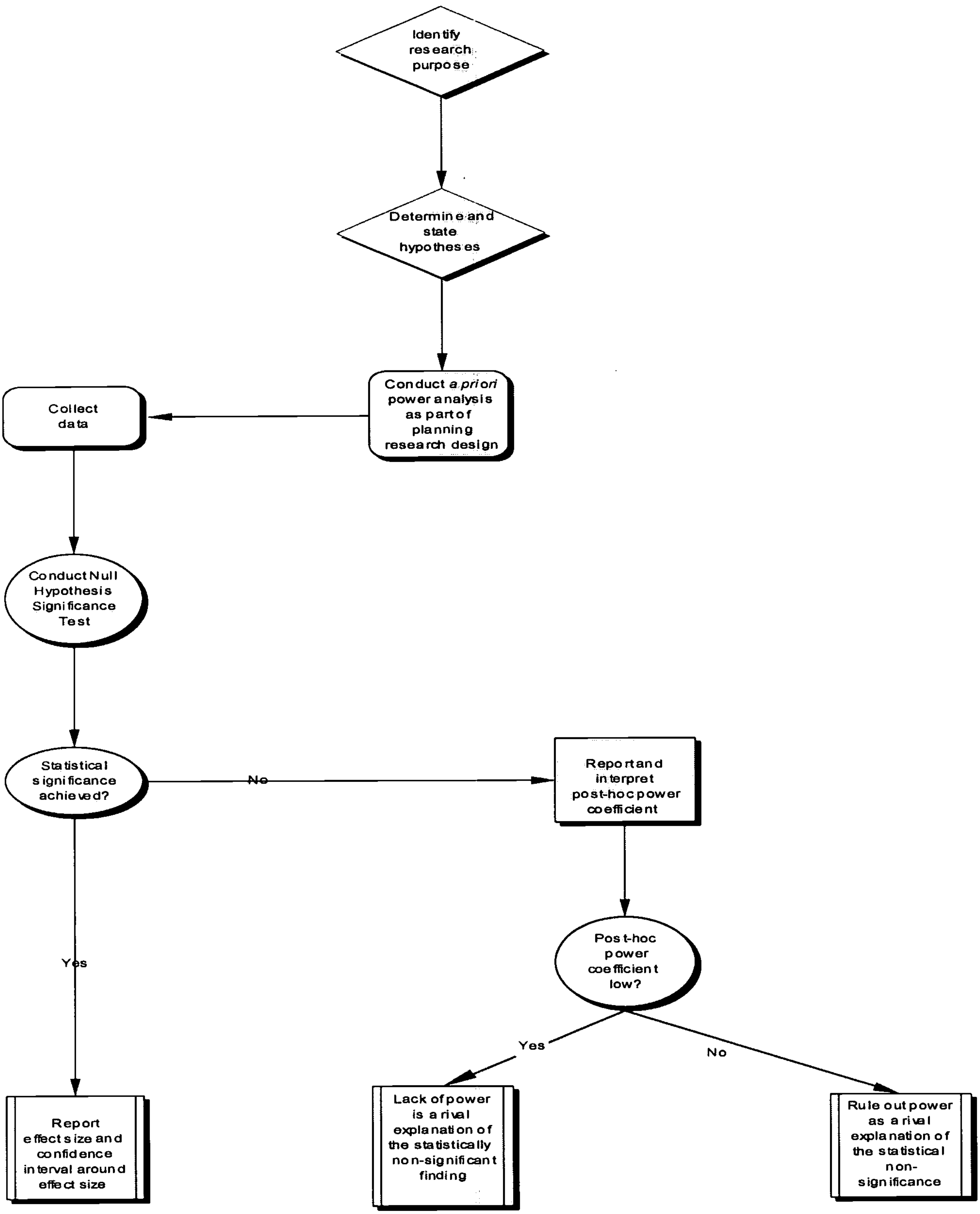
Witcher, A. E., Onwuegbuzie, A. J., & Minor, L. (2001). Characteristics of effective teachers: Perceptions of preservice teachers. *Research in the Schools*, 8, 45-57.

Wolfgang, D. H., & Glickman, C. D. (1986). *Solving discipline problems: Strategies for classroom teachers* (2nd ed.). Boston, MA: Allyn and Bacon.

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

Figure Caption

Figure 1. Power-based Framework for conducting null hypothesis significant tests.



BEST COPY AVAILABLE

Note

[1] Moreover, we recommend that upper bounds for post-hoc power estimates be computed (Steiger & Fouladi, 1997). This upper bound is estimated via the noncentrality parameter. However, this beyond the scope of the present article. For an example of how to compute upper bounds for post-hoc power estimates, the reader is referred to Steiger and Fouladi (1997).

TM034638

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC
Educational Resources Information Center

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Post-Hoc Power: A Concept Whose Time has Come

Author(s): ANTHONY J. ONWUEGBUZIE + NANCY L. LEECH

Corporate Source:

Publication Date: 11/6/02

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☑	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please →

Signature: [signature]

Printed Name/Position/Title: ANTHONY J. ONWUEGBUZIE, Ph.D.

Organization/Address: School of Education, HOWARD UNIVERSITY, 2441 4TH STREET Washington DC, 20059

Telephone: 202-249-0616

FAX: 202-249-0417

E-Mail Address: TONYONWUEGBUZIE@aol.com

Date: 11/7/02

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: info@ericfac.piccard.csc.com
WWW: http://ericfacility.org