**Low Power and Low Reliability:**

**Not a Matter of Concern When the Results are Statistically Significant?**

 My research colleagues and I have too frequently had the experience of having a research manuscript rejected because of low power, even when the results we reported were statistically significant. Here is part of the response I wrote to the editor following one such incident:

 This reviewer finds the reliability (Cronbach alpha = .66) of our misanthropy instrument to be unacceptable, but inexplicably raised no such objection to the similar reliability (.74) of the idealism instrument, the latter being based on twice as many items. J. C. Nunnally (*Psychometric theory* (2nd ed.), 1978, p. 245, New York: McGraw-Hill) recommended that instruments used in basic research (rather than in applied settings, where important decisions may be affected by individual scores) have a reliability of .70 or higher, and added that it would be a waste of effort to attempt to increase reliabilities beyond .80. It is well known that reliability coefficients are greatly influenced by the number of items on a test. Using methods presented in Nunnally (1978, p. 244), I estimated that increasing the number of items on the misanthropy instrument from 5 to 7 or 8 would raise the reliability to approximately .75. Of course, items added would have to be as psychometrically sound as those we already have, and those which we already have are the result of culling items following factor analysis and item analysis of a larger set of items. Please do keep in mind that our current .66 is already awfully close to .70, and it is not at all uncommon for published research reports to include variables based on scales with reliabilities in the .60's (or even with variables based on response to a single Likert-type item). Also, keep in mind that the most likely problem to result from less than desirable reliability in a criterion variable is a reduction in statistical power. Of course, low power is not a matter of concern when the findings are significant – if the findings are statistically significant, then it is clear that the procedures and the analysis used had sufficient power to detect the effect. In fact, one could argue that **significant results when power is not high are much more impressive than significant results when power is high**, since they speak to the detection of an effect of greater magnitude. This is commonly misunderstood, despite the fact that several prominent statisticians have spoken about it. Rosenthal and Gaito (The interpretation of levels of significance by psychological researchers, *Journal of Psychology*, 1963, *55*, 33-38) demonstrated psychologists' lack of understanding of the logic and meaning of tests of statistical significance back in the 60's, and Nelson, Rosenthal, and Rosnow (Interpretation of significance levels and effect sizes by psychological researchers, *American Psychologist*, 1986, *41*, 1299 - 1301) have more recently demonstrated that such misunderstanding remains the norm. While discussing Rosenthal and Gaito's (1963) finding that psychologists place more confidence in studies with large sample sizes (high power), David Bakan (The test of significance in psychological research, *Psychological Bulletin*, 1966, *66*, 423 – 437) stated "the rejection of the null hypothesis when the number of cases is small <that is, power is low> speaks for a more dramatic effect in the population; and if the *p* value is the same, the probability of committing a Type I error remains the same. Thus one can be more confident with <statistically significant results > with a small *n* <low power> than a large *n* <high power>."

 So, what was the editor’s response. He sent my comments to the reviewer. He had described this reviewer as his “methodological expert.” Later he sent the reviewer’s reply back to me. The reviewer wrote “Who the f\*^k does this guy think he is. Tell him to do as I say or just go away.” I just went away, published the manuscript in a better journal, and have not yet submitted any more manuscripts to this editor.

 I should add that even with statistically significant results, low power will result in one's estimation of the effect size having more error – that is, when you construct a confidence interval to show the size of the effect it will be wider than it would have been if you had more power. Accordingly, having more power is not really a liability, as long as you provide interval estimates of effect size and as long as you are not dealing with an audience who thinks that a "significant" effect is a big, important effect – with great power, a significant effect could be a trivial effect.

[Karl L. Wuensch](http://core.ecu.edu/psyc/wuenschk/klw.htm), November, 2006.

[Return to Wuensch’s Stat Help Page](http://core.ecu.edu/psyc/wuenschk/StatHelp/StatHelp.htm)

APA Citation for this Document:

Wuensch, K. L. (2006). *Low power and low reliability: Not a matter of concern when the results are statistically significant?* Retrieved from http://core.ecu.edu/psyc/wuenschk/StatHelp/PowerReliabSignif.docx.