**Nunnally on Reliability**

Nunnally (1978, p. 245) is often associated with the assertion that instruments used in basic research should have reliability of .70 or better. On the other hand, with instruments used in applied settings, a reliability of .80 may not be high enough. Where important decisions about the fate of individuals is made on the basis of test scores, reliability should be at least .90, preferably .95 or better. Here is an excerpt from Nunnally:

*what a satisfactory level of reliability is depends on how a measure is being used. In the early stages of research . . . one saves time and energy by working with instruments that have only modest reliability, for which purpose reliabilities of .70 or higher will suffice. . . . In contrast to the standards in basic research, in many applied settings a reliability of .80 is not nearly high enough. In basic research, the concern is with the size of correlations and with the differences in means for different experimental treatments, for which purposes a reliability of .80 for the different measures is adequate. In many applied problems, a great deal hinges on the exact score made by a person on a test. . . . In such instances it is frightening to think that any measurement error is permitted. Even with a reliability of .90, the standard error of measurement is almost one-third as large as the standard deviation of the test scores. In those applied settings where important decisions are made with respect to specific test scores, a reliability of .90 is the minimum that should be tolerated, and a reliability of .95 should be considered the desirable standard. (pp. 245-246)*

Nunnally (1978, p. 244), shows how to calculate how many additional items one would need to raise the reliability of an instrument to the desired value (assuming that the additional items are as good as the items already on hand).

, where *rd* is the desired reliability , *re* is the reliability of the existing instrument, and *k* is the number of times the test would have to be lengthened to obtained the desired reliability. For example, suppose you have a 5 item test whose reliability is .66. To raise the reliability to .70,  Thus, you would need a test with 1.2(5) = 6 items. To raise the reliability to .75, k = 1.5, you would need 7 or 8 items. To raise the reliability to .80, *k* = 2.06, you would need 10 or 11 items.

If you are stuck with data from an instrument with reliability lower than you desire, can you trust the results? If your results are not significant, I think not. Low reliability increases random error and thus lowers power, making it more likely that your not significant results are a Type II error. If, however, your results are significant, that low power speaks to the correlation found being so large that it produces significant results in spite of the low power due to less than desired reliability. It still would be better to have more power though, as the greater error variance will expand the width of the confidence interval for the effect (lower precision in estimation of the size of the effect).

**Correcting for Measurement Error in Bivariate Linear Correlations**

The following draws upon the material presented in the article by Schmidt and Hunter (1996)

When one is using observed variables to estimate the correlation between the underlying constructs which these observed variables measure, one should correct the correlation between the observed variables for attenuation due to measurement error. Such a correction will give you an estimate of what the correlation is between the two constructs (underlying variables), that is, what the correlation would be if we able to measure the two constructs without measurement error.

Measurement error results in less than perfect values for the reliability of an instrument. To correct for the attenuation resulting from such lack of perfect reliability, one can apply the following correction:

,where

is our estimate for the correlation between the constructs, corrected for attenuation,

*rXY* is the observed correlation between X and Y in our sample,

*rXX* is the reliability of variable X, and

*rYY* is the reliability of variable Y.

Here is an example from my own research:

I obtained the correlation between misanthropy and attitude towards animals for two groups, idealists (for whom I predicted there would be only a weak correlation) and nonidealists (for whom I predicted a stronger correlation). The observed correlation was .02 for the idealists, .36 for the idealists. The reliability (Cronbach alpha) was .91 for the attitude towards animals instrument (which had 28 items) but only .66 for the misanthropy instrument (not surprising, given that it had only 5 items). When we correct the observed correlation for the nonidealists, we obtain , a much more impressive correlation. When we correct the correlation for the idealists, the corrected *r* is only .03.

I should add that Cronbach's alpha underestimates a test's reliability, so this correction is an over-correction. It is preferable to use maximized lamba4 as the estimate of reliability. Using labmda4 estimates of reliability, the corrected *r* is 

**References and Recommended Readings**

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of two commonly reported cutoff criteria. What did they really say? *Organizational Research Methods*, *9*, 202-220.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, *1*, 199-223.

Wuensch, K. L. [Low Power, Low Reliability](http://core.ecu.edu/psyc/wuenschk/StatHelp/PowerReliabSignif.docx) -- Is it a big problem when results are significant?