

Commentary

Commentary on Psychometric concerns in neuropsychological testing

Philip Schatz

*Psychology Department, Saint Joseph's University,
Philadelphia, PA 19131, USA*

Dr. Williams raises interesting concerns in his article, however, I feel that the issues he raises extend beyond the scope of test validation, battery selection, test norms, and standardized assessment. The utilization of an assessment measure involves not only the use of that specific measure within a test battery, but also the validation and normative research for that measure.

The Standards for Educational and Psychological Testing address much needed standards concerning test construction, evaluation, and use [1]. In 1996, the American Psychological Association's Board of Scientific Affairs organized a Task Force on Statistical Inference. Their initial report outlined means of enhancing the quality of data usage in order to protect against potential misrepresentation of quantitative results. They recommended that more extensive descriptions of the data be provided, that researchers characterize analytical results (beyond simple p value statements) to include effect sizes and directions, provide confidence intervals, and assure that reported results are not produced by anomalous data [2].

Utilizing normative or validating studies which do not adhere to these standards may ultimately influence clinical decisions. While published psychometric research is generally sound with respect to methodology and statistical analysis, publications providing ef-

fect sizes and confidence intervals are all too often not the norm. To simply report a Bonferonni correction as a means of guarding against spurious results (when conducting multiple analyses) in the absence of documenting effect sizes provides incomplete data. Ancillary to this issue is the emphasis on the "standard" alpha level of 0.05. There exist excellent overviews of the relationship between p level, sample size, and effect size [3,4]. An alpha level of 0.10 or 0.15 is reasonable for smaller sample sizes [5]; the ultimate result of an empirical study is characterized not by alpha level, but by effect size.

So, while APA's Standards for Educational and Psychological Testing provide much needed guidance for developing and utilizing assessment measures, basing clinical decisions upon studies which do not similarly adhere to APA's Task Force on Statistical Inference may prove counterproductive.

References

- [1] Standards for Educational and Psychological Testing, American Psychological Association, Washington, DC, 1985.
- [2] Task Force on Statistical Inference: Initial Report, American Psychological Association, Washington, DC, 1996, [online] available: <http://www.apa.org/science/tfsi.html>.
- [3] J. Cohen, Things I have learned (so far), *American Psychologist* **45** (1990), 1304–1312.
- [4] R.L. Rosnow and R. Rosenthal, Statistical procedures and the justification of knowledge in psychological science, *American Psychologist* **44** (1989), 1276–1284.
- [5] J. Stevens, *Applied Multivariate Statistics for the Social Sciences*, Lawrence Earlbaum and Associates, Mahwah, NJ, 1996.