The Use of “Effect Size” in Augmenting the Results of Significance Testing: A Comparison of Pre/Post Data from a Geriatric Interclerkship

Urip Purwono, Susan V. Barrett, Michele M. Carlin, Mary L. Zanetti, Sarah McGee
University of Massachusetts Medical School

BACKGROUND
Most research in medical education, when examining the impact of an intervention, report findings based primarily on significance testing despite the controversy of its appropriate use. Moreover, the p-value used to determine rejection or acceptance of the null hypothesis tells nothing about the magnitude of the significance. Using a pre-post assessment of a Geriatric Interclerkship as a case study, this study examines the utility of “effect size” measures in augmenting significance testing results.

Effect size (ES) is a name given to a family of indices that measure the magnitude of a treatment effect. Unlike significance tests, these indices are independent of sample size. The formula used to calculate ES in this study follows:

\[ d = \frac{M_1 - M_2}{\sigma_{\text{pooled}}} \]

\[ \sigma_{\text{pooled}} = \sqrt{\frac{(\sigma_1^2 + \sigma_2^2)}{2}} \]

Where \( M_1 \) and \( M_2 \) are the means of the pre- and post- groups, respectively, and the pooled standard deviation is the square root of the average of the squared standard deviations (Cohen, 1988).

METHOD
A pre-post evaluation consisting of knowledge and attitude items was constructed and administered to students participating in the Geriatric Interclerkship during academic year 2003-2004. The percent of student change on attitude and knowledge items from pre- to post- was measured, matched pre/post data was examined in terms of significance testing (paired t-test) and one measure of “effect size” (Cohen’s d).

RESULTS
Tables 1 and 2 highlight the percent of student pre/post change on attitude and knowledge items. Reliability of the instruments was .63 (knowledge test) and .82 (attitude scale). Ninety and ninety-two matched responses were collected to analyze clusters of attitude and knowledge items, respectively. Table 3 indicates the pre-post difference was significant for both the knowledge and attitude domains (\( t_{90} = 17.60 \) and \( t_{92} = 5.82 \), respectively; \( p < .01 \)). Table 4 shows the effect size for the knowledge domain was very high (\( d = 1.84 \)) indicating a substantial change in student pre to post knowledge, thus complementing the result of the significance testing. On the contrary, the effect size for the attitude domain was small (\( d = .34 \)), indicating a minimal change in student attitude.

CONCLUSION
The “effect size” provides additional practical information to the significance testing. The routine use of this analysis is recommended to enhance the quality of research and evaluation in medical education. More specifically, the results from this Geriatric Interclerkship could eventually be compared to the change in student knowledge and attitude in another Interclerkship experience. Therefore, the most important benefit of reporting effect sizes is that this information provides the researcher with a more standard tool that allows for meta-analysis across studies.