**Defining Effect Sizes**

An effect size is a quantitative indicator of the magnitude of a statistical association. In other words, effect sizes provide information about the magnitude of finding. To be counted in this coding system, the actual effect size must be reported. For example, it is not sufficient to provide a p-value for a Pearson’s r, rather the actual Pearson r value or a mention of its magnitude should be reported to be counted.

Below is a list of common effect sizes that one might encounter in clinical neuropsychology studies. This list is not meant to be exhaustive.

* Pearson’s *r*, Spearman’s *rho*, and Cohen’s q are used characterize the magnitude of an association between two variables. Correlations can be tricky to code as effect sizes. Simply stating that one conducted a correlation or providing a p-value is NOT sufficient to score > 0 in any section. The authors must report an actual correlation coefficient (ranging from -1.0 to 1.0) or mention its size (e.g., small, medium, or large) in order to get a score > 0.
* *R*2, adjusted *R*2, *R*2 delta/change, standardized Beta coefficient, and Cohen’s *f*2 might be used to describe the amount of variance explained in a regression. These statistics are coded in the same manner as correlations, meaning that the authors must report an actual value (e.g., ranging from 0.0 to 1.0) or mention its size (e.g., small, medium, or large) in order to get a score > 0.
* Cohen’s *d*, Cliff’s *d*, eta2, partial eta2, Glass’s delta, omega2 or Hedge’s *g* might be used to describe the magnitude of differences between two or more groups on a continuous variable.
* Cramer’s *V*, Cramer’s Phi, Cohen’s *w*, and Cohen’s *h* might be used to describe the magnitude of associations for categorical variables.
* Classification accuracy statistics can also be effect sizes. These include metrics like area under an ROC curve (AUC), sensitivity, specificity, positive or negative predictive values, odds ratios, risk ratios, hazards ratios, and likelihood ratios. As with correlation coefficients, actual AUC values should be reported in order to qualify as an effect size to warrant a score >1.
* Factor analytic studies may also include measures of effect size. In order to get a score of 3, the authors have to expand upon what these values mean.
	+ For EFA these could include communality or cumulative % of variance explained (most papers will mention the latter). Both of these are akin to an *R*2 effect size. For communality or cumulative % of variance explained >.5 = large effect size or something along those lines.
	+ For CFA these could include RMSEA or the average variance extracted (AVE). RMSEA <0.1= good, <0.08=better, <0.06=best  or the average variance extracted (AVE) ≥ 0.5.

**Effect Size Coding System Scores**

The Abstract, Introduction, Methods, Results, and Discussion section of an article is rated on its reporting and interpretation of effect sizes. Ratings are assigned on a scale ranging from 0 to 3, with higher scores indicating more extensive integration of effect sizes (see below). If you are undecided between two possible scores, choose the lower score.

0 = *Absent (i.e.,* no reporting or mention of effect sizes)

1 = *Minimal* (e.g., A brief mention of the magnitude of an effect in the Introduction or Discussion; A few reliability values in the Methods; or a few correlation coefficients in the Results)

2 = *Moderate* (e.g., Effect sizes are mentioned multiple times or given some interpretive weight in an Introduction or Discussion; Several reliability values in the Methods; or several null hypothesis significance testing analyses in the Results are accompanied by estimates of effect size)

3 = *Extensive* (e.g., A primary focus of the motivation for the study and interpretation of the findings is rooted in effect sizes; Extensive reporting of reliability and validity effect sizes in the Methods; A majority of null hypothesis significance testing analyses are accompanied by an estimate of effect size)

**Coding Criteria for Each Article Section**

Abstract

0 = *Absent*

1 = *Minimal:* Only one section of the abstract includes effect sizes, which might be: A) a brief mention of previous effect sizes in the Introduction; B) stating that effect sizes were conducted in the Methods (again with the correlation caveat); C) reporting a few effect sizes in the Results (i.e., an *actual* correlation coefficient, Cohen’s *d* value, or classification accuracy statistic); or D) mentioning the magnitude of the observed effects in the Conclusions.

2 = *Moderate:* Two or more sections of the abstract includes effect sizes OR one section is extensively based on effect sizes.

3 = *Extensive:* Three or more sections of the abstract include effect sizes or two or more sections are extensively based on effect sizes

Introduction

0 = *Absent*

1 = *Minimal:* A passing mention of the magnitude of an effect or two that is relevant to the study aims (e.g., We tend to see small, negligible effects of MCI on basic ADLs). Actual effect size values are not necessary to warrant a score >0.

2 = *Moderate:* Effect sizes are mentioned multiple times throughout the Introduction and/or there is some integration of the magnitude of findings to set up the study (e.g., We tend to see small, negligible effects of MCI on basic ADLs, but larger effects on instrumental ADLs that might be attributable to MCI’s adverse impact on cognitive domains such as memory that show medium associations with multiple dimensions of everyday functioning).

3 = *Extensive:* Effect sizes are integrated throughout nearly every paragraph in the Introduction, or perhaps serve as the primary factor motivating the study (e.g., a meta-analysis [although of course not all meta-analyses will necessarily warrant a 3 as there is variability in how well this is executed across studies]). Scores of 3 will likely be rare.

Methods (Tables should be considered as Results)

0 = *Absent*

1 = *Minimal:* The authors minimally include one of the following: A) actual statistics for internal consistency (e.g., Cronbach’s alphas) or other aspects of reliability (e.g., test-retest coefficients) or validity (e.g., convergent correlation coefficients) for their study measures; B) A statement indicating that they will include a measure of effect size as part of their Data Analysis section (see caveat above about correlations); C) A power analysis that includes an expected effect size; D) A table of study descriptive variables includes some effect sizes; or E) some other mention of effect sizes per the definitions outlined above.

2 = *Moderate:* The authors include 2 of the above A-E criteria; or the authors include fairly extensive details about 1 of those areas.

3 = *Extensive:* The authors include at least 3 of the above A-E criteria; or a clear majority/thrust of the Methods is about effect sizes.Scores of 3 will likely be rare.

Results (Tables should be considered as Results)

0 = *Absent*

1 = *Minimal:* Between 1-25% of analyses include effect sizes.

2 = *Moderate:* Between 26-75% of analyses include effect sizes.

3 = *Extensive:* More than 75% of analyses include effect sizes.

Discussion

0 = *Absent*

1 = *Minimal:* A passing mention of the magnitude of an effect or two that was observed (e.g., …findings were associated with a small-to-medium effect size). Actual effect size values are not necessary to warrant a score >0.

2 = *Moderate:* Effect sizes are mentioned multiple times throughout the Discussion and/or there is some integration of the magnitude of findings within the study (e.g., we saw small effects X, but large effects on Y) or in relation to the extant literature (e.g., In contrast to the small effects traditionally observed for X, we found a large magnitude association in this study…).

3 = *Extensive:* Effect sizes are integrated throughout nearly every paragraph in the Discussion, or perhaps serve as the primary interpretive factor (e.g., a meta-analysis [although of course not all meta-analyses will necessarily warrant a 3 as there is variability in how well this is executed across studies]). Scores of 3 will likely be rare.