Table 3. Summary of major strengths and weaknesses of each effect size measure reviewed

|  |  |  |
| --- | --- | --- |
| Effect size measure | Strengths | Weaknesses |
| Within-case Standardized Mean (*d*BR) | * Historically, most common method in aphasia & related disorders * Easy to implement | * Influenced by experimental design features (e.g., item set-size, baseline/treatment length) * No clear solution to cases of low baseline variability * Interpretation depends on benchmark study * Lacks a measure of uncertainty |
| Proportion of Potential Maximal Gain  (PMG) | * Accommodates different item set sizes when comparing across individuals * Easy to implement | * Confounded by disorder severity if baseline performance is associated with disorder severity * Can obscure differences in absolute change scores * Lacks a measure of uncertainty |
| Tau-U | * Non-parametric and distribution free * Option to adjust for baseline trends * Easy to implement | * Does not fully characterize the magnitude of change * Influenced by ratio of baseline/treatment observations * Lack of easily interpretable scaling; not bounded between [-1, 1] * Tau-UA VS. B – TREND-A lacks a measure of uncertainty |
| Generalized linear mixed-effects models (GLMM) | * Able to adjust for baseline trends * Effect size available in multiple units of measure (e.g., logits, odds-ratio, percent, items gained), each with clear interpretation * Pools item-level and/or participant-level data to produce more generalizable estimates * Includes confidence/credible interval * Can estimate group and individual effect sizes from a single model (Bayesian) | * Complex to implement * Model convergence challenges are common with frequentist estimation * Confidence/Credible interval width dependent on sample size |