HHS Public Access

Author manuscript

*Arch Phys Med Rehabil*. Author manuscript; available in PMC 2020 June 01.

Published in final edited form as:

*Arch Phys Med Rehabil*. 2019 June ; 100(6): 1131–1139.e87. doi:10.1016/j.apmr.2018.08.177.

# Benchmarks of significant change after post-stroke language rehabilitation

**Natalie Gilmore, MS, CCC-SLP**[1,*], **Michaela Dwyer**[1], and **Swathi Kiran, PhD, CCC-SLP**[1]

[1]Boston University, Sargent College of Health and Rehabilitation Sciences, Speech, Language, and Hearing Sciences, Aphasia Research Laboratory, Room 326, 635 Commonwealth Avenue, Boston, MA, 02215

## Abstract

**Objective:** To establish benchmarks of significant change for aphasia rehabilitation outcome measures (i.e., Western Aphasia Battery-Aphasia Quotient [WAB-AQ], Communicative Effectiveness Index [CETI], Boston Naming Test [BNT]) and assess if those benchmarks significantly differed across subgroups (i.e., time post onset, dose frequency, treatment type).

**Data Sources:** A comprehensive literature search of 12 databases, reference lists of previous reviews, and evidence-based practice materials was conducted.

**Study Selection:** Randomized-controlled trials, quasi-experimental studies, single-subject design, and case studies that used a standardized outcome measure to assess change were included. Titles and full-text articles were screened using a dual review process. 78 studies met criteria for inclusion.

**Data Extraction:** Data were extracted independently and 25% of extractions were checked for reliability. All included studies were assigned quality indicator ratings and an evidence level.

**Data Synthesis:** Random-effects meta-analyses were conducted separately for each study design group (i.e., within/between group comparisons). For within group designs, the summary effect size after aphasia rehabilitation was 5.03 points (95% confidence interval: 3.95-6.10, $p < .001$) on the WAB-AQ, 10.37 points (6.08-14.66, $p < .001$) on the CETI and 3.30 points (2.43-4.18, $p < .001$) on the BNT. For between group designs, the summary effect size was 5.05 points (1.64-8.46, $p = .004$) on the WAB-AQ, and .55 points (−1.33, 2.43, $p = .564$) on the BNT, the latter of which was not significant. Subgroup analyses for the within group designs showed no significant differences in the summary effect size as a function of dose frequency, or treatment type.

**Conclusions:** This study established benchmarks of significant change on three standardized outcome measures used in aphasia rehabilitation.

### Keywords

stroke; rehabilitation; outcome; speech therapy; aphasia

[*]Corresponding author: Phone: 617-353-2706 Fax: 617-353-5074 ngilmore@bu.edu.

Thirty to forty percent of stroke survivors experience aphasia.[1] While numerous systematic reviews and meta-analyses have demonstrated aphasia rehabilitation efficacy,[2,3] none have provided the average significant change, or summary effect size (ES) by outcome measure, a valuable metric for practitioners and researchers. Robey's hallmark meta-analyses[2,4,5] showed a positive aphasia treatment effect, but were segregated by study design and focused on identifying the effect size for different conditions (e.g., treated vs untreated recovery). Similarly, the most recent Cochrane review[3] demonstrating speech therapy efficacy, synthesized data from randomized controlled trials only, excluding a wealth of aphasia treatment data. Furthermore, effect sizes were represented as standardized mean differences for specific behaviors (e.g., verbal expression), not for specific outcome measures (e.g., Western Aphasia Battery-Aphasia Quotient[6] [WAB-AQ]).

Another option is to synthesize results by outcome measure to obtain a summary ES (i.e., raw unstandardized mean difference),[7] which can be used to interpret meaningful change on a specific assessment post-treatment. Clinicians and researchers frequently utilize standard error of measurement (SEM) to interpret a test score's meaningfulness after intervention. However, summary ES is a more appropriate metric. It reflects the treatment effect's size[7] and can be used to interpret group data, as opposed to SEM, which is more relevant for interpreting individual scores.[8]

Numerous aphasia assessment instruments exist[9] for assessing impairment (i.e., Body Structure/Function), functional communication (i.e., Activity/Participation), psychosocial functioning (i.e., Contextual Factors) and well-being (i.e., Quality of Life [QOL]). It is not surprising then that practicing speech-language pathologists[10–12] and researchers[13,14] use measures inconsistently making synthesis and comparison across trials challenging.

Wallace and colleagues proposed a core outcome set (COS)[13,15–18] for aphasia, specifying a minimum set of outcomes that should be administered to persons with aphasia as standard practice (i.e., WAB, The Scenario Test, General Health Questionnaire-12, SAQOL-39g) to increase consistency. Yet, the summary ES for these measures remains unknown. Given the potential benefits to clinical and research practice, a systematic review of behavioral aphasia intervention studies with meta-analyses was conducted with two aims: 1) To calculate the summary ES reported on the most frequently-used and relevant outcome measures; and 2) To determine if the summary ES significantly differed across subgroups for each outcome measure (i.e., time post onset, dose frequency, treatment type).

## METHODS

This study followed the Preferred Reporting Items for Systematic Review and Meta-analyses: the PRISMA Statement[19] guidelines and was registered at the International prospective register of systematic reviews, PROSPERO, under the identification number CRD42016039393.

### Inclusionary Criteria

Randomized-controlled, quasi-experimental, single-subject design, and case studies with an $n \geq 3$ were included if they (1) assessed the effect of a behavioral aphasia intervention and

(2) used a standardized outcome measure to evaluate change post-treatment as compared to pre-treatment (i.e., data from two time points).

## Literature search

The following databases: PubMed, EMBASE, CINAHL, PsycINFO, SpeechBite, LLBA, PLoS, Worldcat, Web of Science, Ageline, Scopus, and Google Scholar were searched (see Supplementary Material 1 for sample search strategy) from 5/24/2016-08/26/2016. Reference lists of relevant systematic reviews, meta-analyses and professional organization materials were reviewed. Search terms were modified to meet each database's requirements. Grey literature was removed during screening. All citations were managed using Zotero[20] and exported to Excel for screening and data extraction.

## Study Selection and Data Extraction

Two reviewers (first two authors) independently screened 9,285 titles and abstracts against inclusionary criteria (96% inter-reviewer reliability). Full-text articles were obtained for records that met all criteria. Both reviewers screened 858 full-text articles against the inclusionary criteria (90% inter-reviewer reliability). Disagreements were resolved through discussion and searching the full-text. Study exclusion rationale was documented (Figure 1). When results from the same dataset were included in multiple publications, only the publication with the greatest sample size was included. Both reviewers extracted the following data from the full-text: the standardized outcome measure used to measure intervention-related change, presence/absence of data from two time points, study design, sample size, testing time points, and population treated (i.e., stroke survivors and/or caregivers).

The number of studies using each standardized outcome measure was calculated. Based on the measure's use frequency (Supplementary Material 2), field relevance (i.e., part of aphasia COS), and disability domain [21,22] measured (i.e., Body Structure/Function, Activity/Participation, Contextual Factors and/or QOL), the WAB-AQ, the Communicative Effectiveness Index[23] (CETI) and the Boston Naming Test[24] (BNT) were chosen for meta-analysis. To have a power of .80 to detect an effect size of ≥.50 using a random-effects model, outcome measures with cumulative sample sizes across within group studies < than 100 were excluded and/or if the measure was used in less than < 10 studies.[25] The contextual factor and QOL COS measures were excluded from meta-analysis because 1) the 12-item General Health Questionnaire was only used in 1 study and 2) sensitivity to change had already been established[26,27] for the Stroke and Aphasia Quality of Life Scale-39. 78 studies met eligibility for meta-analysis. Both reviewers extracted the following data from these studies: age, sex, aphasia type and severity, time post onset, treatment type and description, session length, weekly session frequency, testing time points, treatment length, pre- and post-treatment test score correlation, and pre- and post-treatment mean (SD) on the WAB-AQ, CETI and/or BNT.

Studies were classified as including an acute (i.e., < 6 months post stroke onset) or chronic sample; providing a lower dose frequency (i.e., ≤ 4 hours/week) or a higher dose frequency; and utilizing an impairment-based (i.e., treated discrete deficits), activity/participation-based

(i.e., targeted everyday communication) and/or integrated (i.e. combined impairment and activity/participation level approaches) treatment. According to Warren, Fey and Yoder, 2007,[28] dose frequency is the number of times an intervention was provided daily and weekly.

The same two reviewers responsible for screening divided the data extraction. Each reviewer extracted data for 25% of the others' studies (98% inter-reviewer reliability). Reviewers contacted original authors for additional data needed to calculate effect sizes as needed.

## Quality Assessment

The same two reviewers independently appraised included studies' quality using indicators identified by the American Speech-Language Hearing Association (ASHA) level of evidence scheme.[29,30] See Supplementary Material 3 for quality indicator details. Quality indicator summative scores ≤ 1 for within group studies [Post-treatment Mean vs. Pre-treatment mean for the same group] and ≤ 2 for between group studies [Experimental group Post-treatment Change vs. Control group Post-treatment Change] were excluded for poor quality. Reviewers assigned each study's evidence level using ASHA[31] guidelines originally proposed by the Scottish Intercollegiate Guidelines Network[32] (i.e., IB: randomized controlled study; IIA: non-randomized controlled study; IIB: quasi-experimental study; III: non-experimental studies).

## Data Analysis

Individual patient results from studies with sample sizes ≥ three were averaged to calculate a group mean and SD. Pre-post treatment correlation scores were calculated for studies providing individual subject data as follows: Pre-treatment SD + Post-treatment SD – Change SD/ 2 * Pre-treatment SD * Post-treatment SD.[33] When it could not be computed, the average of the observed pre-post treatment correlation coefficients was used.[34] For crossover designs, data were extracted after both treatment phases, as long as both involved the same treatment type (i.e., impairment, activity/participation and/or integrated). For the WAB-AQ within group analysis, a weighted mean and SD was calculated for the Cherney, 2010 study as the published results were split by severity and for the Mozeiko et al., 2016 study, data for the higher dose frequency and lower dose frequency groups were entered separately.

Meta-analyses were conducted independently for within and between group study designs to avoid methodological concerns involved in transforming to a common metric.[35] After group averages were calculated for both time points, single-subject design and case study data were included in the within group meta-analyses.

Meta-analyses for each outcome measure for both study designs were performed using Comprehensive Meta-Analysis software.[36] As heterogeneity between studies was anticipated, a random-effects model was used to combine individual study results into a summary ES (i.e., raw unstandardized mean difference). Raw unstandardized mean difference was calculated because clinicians and researchers interpret raw change on these outcome measures post-intervention, making this effect size inherently meaningful to the field.[7] Q and $I^2$ statistics were examined to determine the extent of any remaining

heterogeneity across studies. Even if the heterogeneity was low (i.e., non-significant and < 75%), subgroup analyses were conducted to assess summary ES differences depending on recovery stage, treatment type, and dose frequency. Sub-group analyses were corrected for multiple comparisons using the Bonferroni correction method.

### Subgroup Analyses

Although no significant heterogeneity was present in the overall summary ESs, subgroup analyses were performed to investigate for summary ES differences due to these variables. As > 5 studies per subgroup are required to conduct a valid subgroup analysis,[7] the same subgroup analyses were not feasible for all outcome measures and study design groups. Subgroup analyses were conducted with the following variables, outcome measures, and study designs: 1) dose frequency for within group studies using the WAB-AQ, CETI, and BNT and 2) treatment type for within group studies using the WAB-AQ and BNT. No subgroup analyses were conducted to assess for differences in summary ES related to TPO as the nearly all of the within group studies included participants in the chronic phase. No subgroup analysis was conducted to assess for a difference in summary ES according to treatment type for within group studies using the CETI, or any of the between group study designs as there were < 5 studies in each subgroup.

Funnel plots for meta-analyses including > 10 studies were examined for asymmetry (i.e., within group meta-analyses only). Publication bias was objectively assessed using Begg and Mazumdar rank correlation, Egger's regression intercept and Duval and Tweedie's Trim and Fill.[7]

## RESULTS

### Aim 1: What is the summary ES post-therapy on three commonly-used outcome measures in aphasia rehabilitation?

**Study Identification/Description.—**78 studies met criteria for inclusion in the meta-analyses (i.e., within group: 70; between group: 8). Descriptive information and references for these studies can be found in Supplementary Materials 4 through 9.

**Within group study designs.—**Combining individual studies' findings resulted in a significant summary ES indicating a positive treatment effect across all three outcome measures. On the WAB-AQ (53 studies, $n = 522$), the summary ES on the raw unstandardized mean difference was 5.03 points, (95% confidence interval [CI]: 3.95-6.10, $p < .001$). No significant heterogeneity was found (Q = 50.79, df = 52, $p = .52$; $I^2 = 0$). The CETI summary ES (17 studies, $n = 208$), was 10.37 points (6.08-14.66, $p < .001$). No significant heterogeneity was found (Q = 16.47, df = 16, $p = .42$; $I^2 = 2.86$). The summary ES for the BNT (36 studies, $n = 347$), was 3.30 points (2.43-4.18, $p < .001$). No significant heterogeneity was found (Q = 42.17; df = 35; $p = .19$; $I^2 = 17.01$). See Figures 2 and 3 for forest plots depicting the variability across studies.

**Publication bias for within group meta-analyses.—**No marked asymmetry was noted in funnel plots for any of these meta-analyses (Supplementary Materials 10). For the

WAB-AQ, both the Egger's regression intercept ($B$ = 1.31, CI = (−.11, 2.72), $t$ (51) =1.86, $p$ = .04) and the Duval and Tweedie's Trim and Fill (Observed point estimate = 5.03(3.95, 6.10); Imputed point estimate = 5.88 (4.74, 7.02)) suggested the presence of publication bias for the WAB-AQ (i.e., missing positive studies). There was no significant presence of publication bias for the CETI meta-analysis (1-tailed p > .05). For the BNT, the Duval and Tweedie's Trim and Fill revealed the presence of publication bias (Observed point estimate = 3.30(2.43, 4.18); Imputed point estimate = 2.97(2.02, 3.92)) (i.e., missing negative studies). In both cases where publication bias, was indicated, the SES shifted only minimally (i.e., < 1 point, within the confidence interval), verifying that the within group SESs reported for all three outcome measures are valid and can be utilized with confidence.

**Between group study designs.—**On the WAB-AQ (6 studies, Experimental $n$ = 119; Control $n$ = 99), the summary ES on the raw unstandardized mean difference between the experimental and control groups was 5.05 (1.64-8.46, $p$ < .01). No significant heterogeneity was found (Q = 5.26, df = 5, $p$ =.39; $I^2$ = 4.87). No between-group meta-analysis was conducted for the CETI as only one publication using it to measure post-intervention change was identified. On the BNT (5 studies, Experimental $n$ = 66; Control $n$ = 35), the raw unstandardized mean difference between the experimental and control groups at post-treatment was .55 (−1.33-2.43, $p$ = .56). There was no significant heterogeneity between included studies (Q = .86, df = 4, $p$ = .93; $I^2$ = 0). See Figure 4 for forest plots that illustrates the variability across studies.

**Publication bias for between group meta-analyses.—**Due to the low sample size in the between group study design meta-analyses,[37] funnel plots could not be validly assessed for the presence of publication bias.

### Aim 2: Does the summary ES vary according to time post onset, dose frequency and/or treatment type?

There were no statistically significant differences between summary ESs for any of the within group study design subgroup analyses completed (i.e., dose frequency for WAB-AQ, CETI, and BNT; treatment type for WAB-AQ and BNT). See Table 1 for results and Supplementary Materials 11 for forest plots.

### Quality Appraisal

For within group study designs, 73% of studies included in the meta-analyses were level III evidence,[29,31] 26% were IIB, and 1% were IIA. For between group study designs, 50% were classified as IB, 38% as IIA, and 13% as IIB level evidence. None of the 78 studies selected for meta-analysis were excluded from the analysis based on their quality, which is unsurprising as studies of poorer quality were likely excluded during the two initial screening phases. See Table 2 for summative quality indicator scores for both study designs. For within group studies, most studies had summative scores of 3, with higher scores indicating better quality. For between groups comparisons, the majority of studies using the WAB or BNT had summative scores of 7 or 5, respectively. Individual study ratings are included in Supplementary Materials 4-8. The percentage of studies meeting criterion for each specific quality indicator are available in Supplementary Material 12.

## DISCUSSION

This study established benchmarks for significant change on three outcome measures used in aphasia rehabilitation to assess severity, functional communication, and naming ability. Practitioners can use these metrics to objectively demonstrate improvement in their clients following treatment, an essential element of clinical practice that directly influences reimbursement and clients' duration of services. Likewise, researchers can reference the reported summary ESs when quantifying change from experimental interventions, but also when conducting *a priori* power analyses for future studies. The latter analyses require estimating the effect size,[38] which is not consistently reported in published aphasia treatment studies,[39] further emphasizing the utility of this study's benchmarks.

The relationship between the summary ESs established in this study and each outcome measure's SEM must be discussed. WAB-AQ summary ESs (Within group: 5.03; Between group: 5.05), were equivalent to its SEM of 5, which has been framed as a metric of clinically meaningful improvement.[40–42] On initial inspection, the adjacency of these two values suggests a diminished effect of aphasia rehabilitation as measured by the WAB-AQ. However, the seminal work of Hula, Donovan, Kendall & Gonzalez-Rothi, 2010,[42] demonstrating that the WAB-AQ's SEM was actually closer to 2 for AQs between 28-68, but much higher (i.e., up to 12) for scores outside that range (i.e., AQs of 0-27, 69-100) serves to clearly distinguish the summary ES established in this study from measurement error. Future research should examine how the WAB-AQ summary ES varies for persons with more mild or severe aphasia and examine which treatment approaches result in summary ESs well outside of the SEM for all severity groups. The CETI's summary ES of 10.37 was well above its SEM of 5.87,[23] suggesting that those improvements were not due to variations inherent to measurement alone. Lastly, the summary ES for the BNT of 3.30 was also higher than its SEM of 2.04,[43] supporting its validity as a metric of intervention-related improvement. Importantly, the summary ESs were consistent across treatment approaches and dose frequencies as none of the meta-analyses demonstrated significant heterogeneity, nor were any of the sub-group analyses significant.

This study provides a unique contribution to the literature on aphasia rehabilitation as it included studies according to the outcome measure used to assess change as opposed to by study design, as in previous systematic reviews and meta-analyses.[2,3] This methodological shift is valuable as rather than conducting only meta-analyses with between group comparisons, separate meta-analyses were also conducted using within group study comparisons, including single subject design studies. This approach allowed for the inclusion and synthesis of a larger body of the treatment literature in the field than previous reviews. In summary, this work adds to the body of literature that confirms a positive effect of aphasia treatment and further, provides benchmarks for significant change.

Nonetheless, some open questions remain. In order to maintain adequate power to conduct meta-analyses, a number of studies employing less-frequently used outcome measures were excluded (e.g., assessing contextual factors). Secondly, subgroup analyses could not be conducted between acute and chronic participant studies. Third, as the summary ES for the WAB-AQ was only notably higher than the SEM for a range of AQs (i.e., 28-68), it should

be tested whether a higher benchmark for improvement should be used for individuals who are more mild or severe, or a different assessment measure altogether.

### Study Limitations

All systematic reviews and meta-analyses are susceptible to publication bias. Although funnel plots for the within group designs were largely symmetric, publication bias was detected in the within-group WAB-AQ and BNT analyses. However, the point estimates varied minimally and thus, the observed summary ESs for those measures should be considered valid.

## CONCLUSIONS

By combining evidence from existing treatment studies, the present systematic review and meta-analyses establishes valuable benchmarks of change for three frequently used outcome measures. Furthermore, it confirms that aphasia rehabilitation is indeed effective.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was supported by NIH/NIDCD grant T32DC0130170. None of the authors of this work has a financial conflict of interest with respect to this project.

## ABBREVIATIONS

| | |
|---|---|
| **ASHA** | American Speech-Language Hearing Association |
| **BNT** | Boston Naming Test |
| **CI** | Confidence interval |
| **CETI** | Communicative Effectiveness Index |
| **COS** | Core Outcome Set |
| **ES** | effect size |
| **PRISMA** | Preferred Reporting Items for Systematic Review and Meta-analyses |
| **QOL** | Quality of Life |
| **SEM** | Standard Error of Measurement |
| **TPO** | time post onset |
| **WAB-AQ** | Western Aphasia Battery-Aphasia Quotient |

## REFERENCES


1. Dickey L, Kagan A, Lindsay MP, Fang J, Rowland A, Black S. Incidence and Profile of Inpatient Stroke-Induced Aphasia in Ontario, Canada. Arch Phys Med Rehabil. 2010;91(2):196–202. doi: 10.1016/j.apmr.2009.09.020 [PubMed: 20159121]
2. Robey RR. A meta-analysis of clinical outcomes in the treatment of aphasia. J Speech Lang Hear Res. 1998;41(1):172–187. [PubMed: 9493743]
3. Brady MC, Kelly H, Godwin J, Enderby P, Campbell P. Speech and language therapy for aphasia following stroke In: The Cochrane Collaboration, ed. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd; 2016 http://doi.wiley.com/10.1002/14651858.CD000425.pub4. Accessed July 22, 2016.
4. Robey R The Efficacy of Treatment for Aphasia Persons: A Meta-analysis. Brain Lang. 1994;47:582–608. [PubMed: 7859056]
5. Robey R, Schultz M, Crawford A, Sinner C. Review: Single-subject clinical-outcome research: designs, data, effect sizes, and analyses. Aphasiology. 1999;13(6):445–473.
6. Kertesz A. Western Aphasia Battery (Revised). San Antonio, TX: PsychCorp; 2006.
7. Borenstein M, ed. Introduction to Meta-Analysis. Chichester, U.K: John Wiley & Sons; 2009.
8. Harvill LM. Standard error of measurement. Educ Meas Issues Pract. 1991; 10(2) :33–41.
9. Wallace SJ, Worrall L, Rose T, Le Dorze G. Measuring outcomes in aphasia research: A Review of current practice and an agenda for standardisation. Aphasiology. 2014;28(11):1364–1384. doi: 10.1080/02687038.2014.930262
10. Hesketh A, Hopcutt B. Outcome measures for aphasia therapy: It's not what you do, it's the way that you measure it. Eur J Disord Commun. 1997;32(3, Spec Iss):189–202. doi: 10.1080/13682829709177096
11. Worrall L, Egan J. A survey of outcome measures used by Australian speech pathologists. Asia Pac J Speech Lang Hear. 2001;6(3):149–162. doi:10.1179/136132801805576635
12. Simmons-Mackie N, Threats TT, Kagan A. Outcome assessment in aphasia: A survey. J Commun Disord. 2005;38(1):1–27. doi:10.1016/j.jcomdis.2004.03.007 [PubMed: 15475012]
13. Wallace SJ, Worrall L, Rose T, Le Dorze G. Measuring outcomes in aphasia research: A Review of current practice and an agenda for standardisation. Aphasiology. 2014;28(11): 1364–1384. doi: 10.1080/02687038.2014.930262
14. Ali M, English C, Bernhardt J, Sunnerhagen KS, Brady M, VISTA-Rehab Collaboration. More outcomes than trials: a call for consistent data collection across stroke rehabilitation trials: Review. Int J Stroke. 2013;8(1):18–24. doi:10.1111/j.1747-4949.2012.00973.x [PubMed: 23280265]
15. Wallace SJ, Worrall L, Rose T, et al. Which outcomes are most important to people with aphasia and their families? an international nominal group technique study framed within the ICF. Disabil Rehabil. 6 2016:1–16. doi:10.1080/09638288.2016.1194899
16. Wallace SJ, Worrall L, Rose T, Le Dorze G. Which treatment outcomes are most important to aphasia clinicians and managers? An international e-Delphi consensus study. Aphasiology. 5 2016:1–31. doi:10.1080/02687038.2016.1186265
17. Wallace SJ, Worrall L, Rose T, Le Dorze G. Core Outcomes in Aphasia Treatment Research: An e-Delphi Consensus Study of International Aphasia Researchers. Am J Speech Lang Pathol. 2016;25(4S):S729. doi:10.1044/2016_AJSLP-15-0150 [PubMed: 27997949]
18. Wallace S, Worrall L, Rose T, Le Dorze G. Improving research outcome measurement in aphasia: Development of a core outcome set. Presented at the: International Aphasia Rehabilitation Conference (IARC); December 14, 2016; London.
19. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P). Syst Rev. 2015;4(1):1–9. [PubMed: 25554246]
20. Center for History and New Media. Zotero Quick Start Guide. http://zotero.org/support/quick_start_guide.
21. Kostanjsek N Use of The International Classification of Functioning, Disability and Health (ICF) as a conceptual framework and common language for disability statistics and health information systems. BMC Public Health. 2011;11(Suppl 4):S3. doi:10.1186/1471-2458-11-S4-S3

22. Kagan A, Simmons-Mackie N, Rowland A, et al. Counting what counts: A framework for capturing real-life outcomes of aphasia intervention. Aphasiology. 2008;22(3):258–280.
23. Lomas J, Pickard L, Bester S, Elbard H, Finlayson A, Zoghaib C. The Communicative Effectiveness Index: Development and Psychometric Evaluation of a Functional Communication Measure for Adult Aphasia. J Speech Hear Disord. 1989;54(1): 113–124. doi :10.1044/jshd.5401.113 [PubMed: 2464719]
24. Goodglass H, Kaplan E, Weintraub S. Boston Naming Test. Philadelphia: Lea & Febiger; 1983.
25. Liu J Statistical Power in Meta-Analysis. 2015 https://scholarcommons.sc.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=4230&context=etd.
26. Hilari K, Lamping DL, Smith SC, Northcott S, Lamb A, Marshall J. Psychometric properties of the Stroke and Aphasia Quality of Life Scale (SAQOL-39) in a generic stroke population. Clin Rehabil. 2009;23(6):544–557. doi:10.1177/0269215508101729 [PubMed: 19447841]
27. Guo YE, Togher LE, Power E, et al. Sensitivity to change and responsiveness of the Stroke and Aphasia Quality-of-Life Scale (SAQOL) in a Singapore stroke population. Aphasiology. 2017;31(4):427–446.
28. Warren SF, Fey ME, Yoder PJ. Differential treatment intensity research: A missing link to creating optimally effective communication interventions. Ment Retard Dev Disabil Res Rev. 2007;13(1):70–77. doi:10.1002/mrdd.20139 [PubMed: 17326112]
29. Faroqi-Shah Y, Frymark T, Mullen R, Wang B. Effect of treatment for bilingual individuals with aphasia: A systematic review of the evidence. NEL J Neurolinguistics. 2010;23(4):319–341.
30. Mullen R The State of the Evidence: ASHA Develops Levels of Evidence for Communication Sciences and Disorders. ASHA Lead. 2007;12(3):8–25.
31. Steps in the Process of Evidence-Based Practice: Assessing the Evidence. American Speech-Language-Hearing Association. http://www.asha.org/Research/EBP/Assessing-the-Evidence/.Accessed October 4, 2017.
32. SIGN 50: A guideline developer's handbook - SIGN grading system 1999 – 2012. http://www.sign.ac.uk/guidelines/fulltext/50/annexoldb.html. Accessed October 4, 2017.
33. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Vol 4 England: John Wiley & Sons Ltd.; 2011.
34. Laird KT, Tanner-Smith EE, Russell AC, Hollon SD, Walker LS. Short-term and Long-term Efficacy of Psychological Therapies for Irritable Bowel Syndrome: A Systematic Review and Meta-analysis. Clin Gastroenterol Hepatol. 2016;14(7):937–947.e4. doi:10.1016/j.cgh.2015.11.020 [PubMed: 26721342]
35. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychol Methods. 2002;7(1):105–125. doi: 10.1037//1082-989X.7.1.105 [PubMed: 11928886]
36. Borenstein M, Hedges L, Higgins J, Rothstein H. Comprehensive Meta-Analysis. Englewood, NJ: Biostat; 2014.
37. Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343(7221):d4002–d4002. doi:10.1136/bmj.d4002 [PubMed: 21784880]
38. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. 3rd Edition. Upper Saddle River, NJ: Prentice Hall Health; 2000. 2008.
39. Beeson PM, Robey RR. Evaluating Single-Subject Treatment Research: Lessons Learned from the Aphasia Literature. Neuropsychol Rev. 2006;16(4):161–169. doi:10.1007/s11065-006-9013-7 [PubMed: 17151940]
40. Elman RJ, Bernstein-Ellis E. The efficacy of group communication treatment in adults with chronic aphasia. J Speech Lang Hear Res. 1999;42(2):411–419. [PubMed: 10229456]
41. Katz RC, Wertz RT. The efficacy of computer-provided reading treatment for chronic aphasic adults. J Speech Hear Res. 1997;40(3):493–507.
42. Hula W, Donovan NJ, Kendall DL, Gonzalez-Rothi LJ. Item response theory analysis of the Western Aphasia Battery. Aphasiology. 2010;24(11): 1326–1341. doi: 10.1080/02687030903422502

43. Flanagan JL, Jackson ST. Test-Retest Reliability of Three Aphasia Tests: Performance of Non-Brain-Damaged Older Adults. J Commun Disord. 1997;30(1):33–43. [PubMed: 9017477]

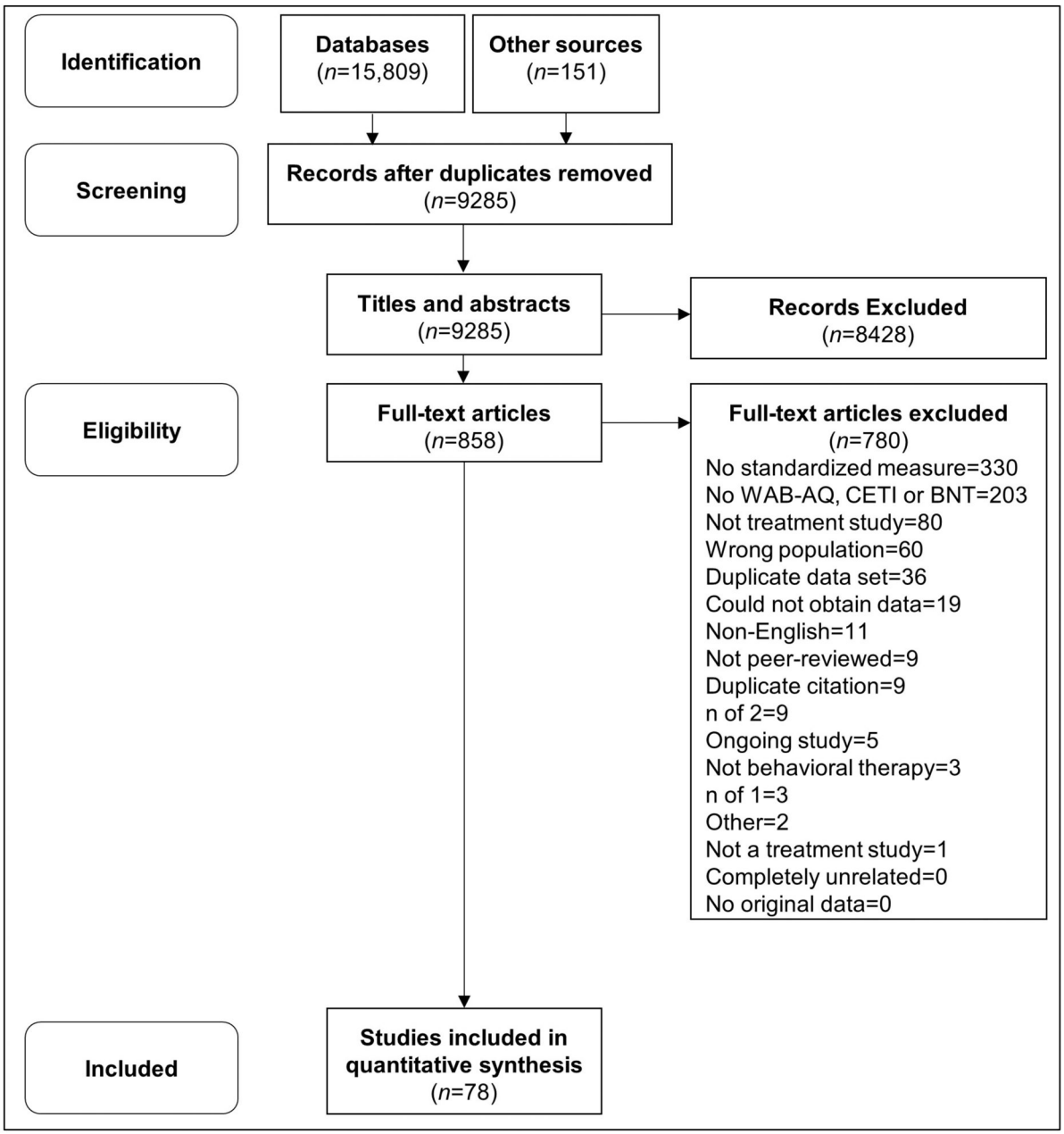


**Figure 1.**

The PRISMA flow diagram1 of study inclusion. *Note:* 1. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses The PRISMA Statement. *PLoS Med*. 2009;6(7):6.

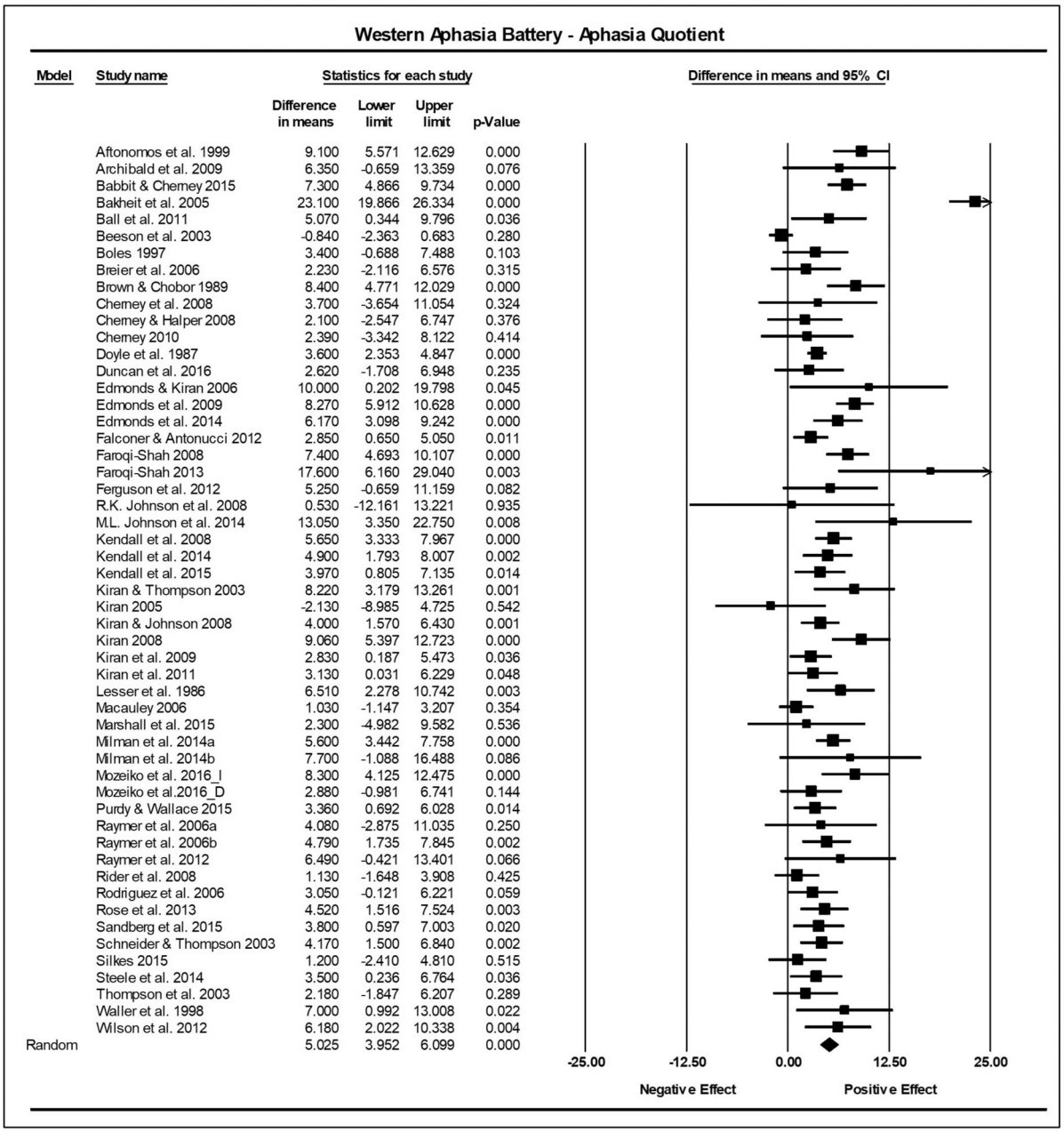


**Figure 2.**

Summary effect sizes for within group studies reporting the Western Aphasia Battery-Aphasia Quotient (WAB-AQ). The difference in means column reflects the pre-treatment mean subtracted from the post-treatment mean. The lower and upper limits columns show the 95% confidence interval surrounding the difference in means. The p-value indicates the significance of the effect. The final row describes the summary effect size, 95% confidence interval, and p-value. The diamond represents the summary effect size. The squares reflect effect sizes of individual studies.

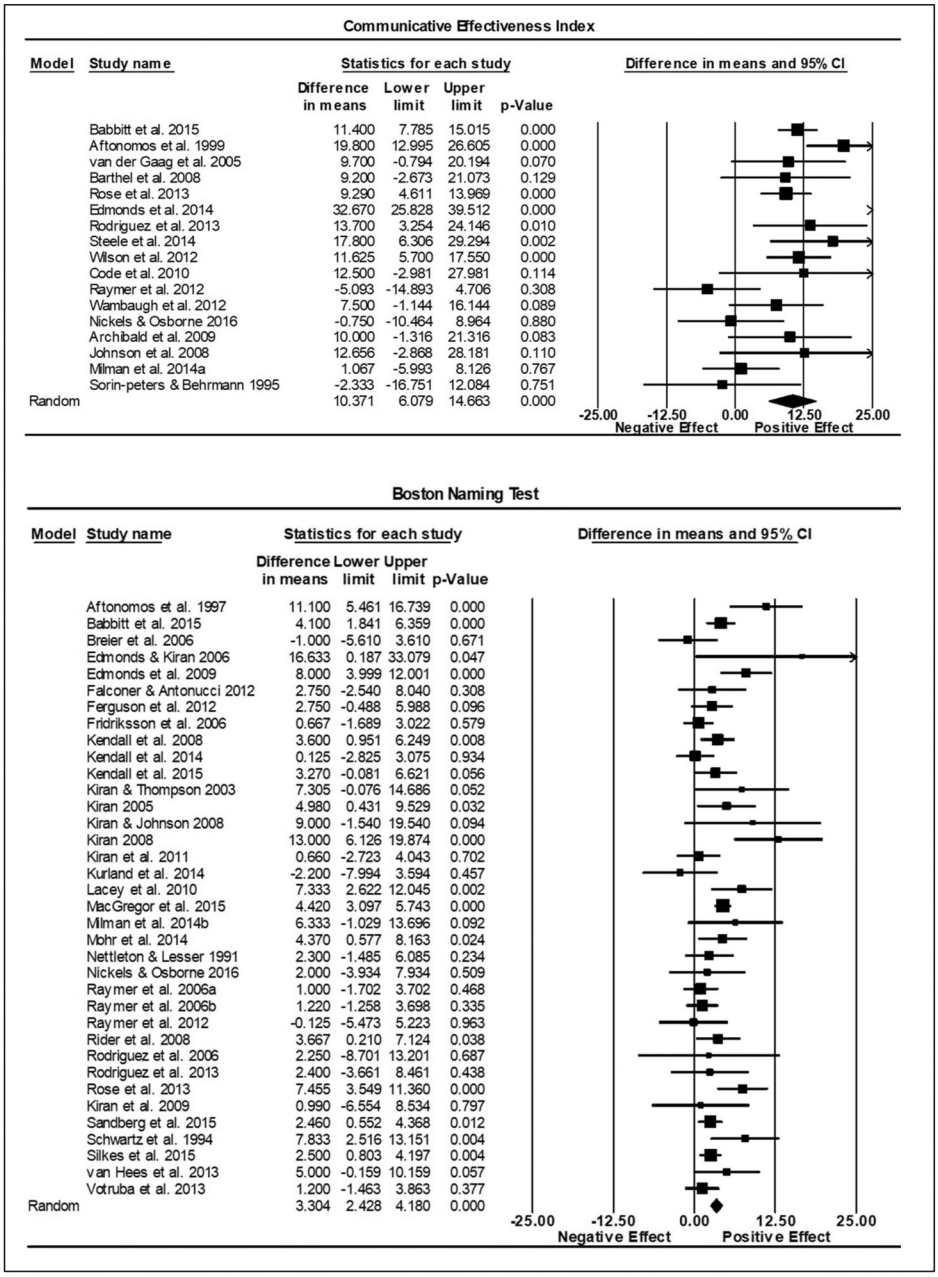


**Figure 3.**

Summary effect sizes for within group studies reporting the Communicative Effectiveness Index (CETI) and Boston Naming Test (BNT). Figure details are the same as for Figure 2.

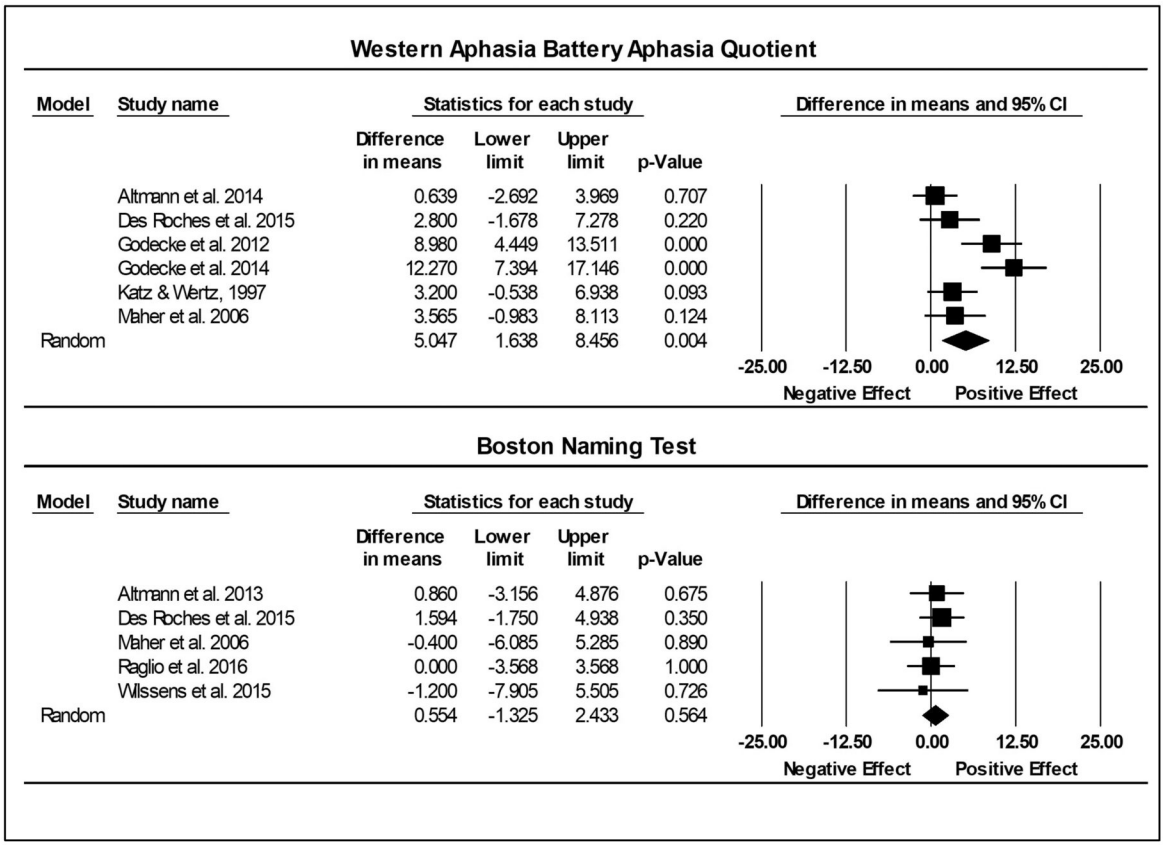


**Figure 4.**

Summary effect sizes for between group studies reporting the Western Aphasia Battery-Aphasia Quotient (WAB-AQ) and Boston Naming Test (BNT). The diamond is the summary effect size. The squares reflect effect sizes of individual studies. The difference in means column reflects the post-treatment control group mean change subtracted from the post-treatment experimental group mean change. The lower and upper limits columns show the 95% confidence interval surrounding the difference in mean change. The p-value indicates the significance of the effect. The final row describes the summary effect size, 95% confidence interval, and p-value. The diamond represents the summary effect size. The squares reflect effect sizes of individual studies.

**Table 1.**

Results of subgroup analyses for within group study designs

| Outcome Measure | LDF | HDF | IMP | A/P | INT |
|---|---|---|---|---|---|
| | *n* = 35 | *n* = 11 | *n* =33 | *n* = 6 | *n* = 14 |
| WAB-AQ | | | | | |
| | 4.50 | 5.17 | 4.42 | 5.10 | 6.48 |
| | 3.64-5.36 | 3.72-6.61 | 3.09-5.76 | 1.73-8.47 | 4.38-8.57 |
| | *n* = 10 | *n* = 5 | | | |
| CETI | | | | | |
| | 10.05 | 11.02 | n/a | n/a | n/a |
| | 3.83-16.28 | 2.81-19.24 | | | |
| | *n* = 25 | *n* = 9 | *n* = 24 | *n* = 5 | *n* = 7 |
| BNT | | | | | |
| | 3.55 | 3.39 | 3.18 | 3.89 | 3.34 |
| | 2.33-4.76 | 1.75-5.02 | 2.09-4.27 | 1.65-6.14 | 1.18-5.49 |

Note: WAB-AQ=Western Aphasia Battery-Aphasia Quotient; CETI= Communicative Effectiveness Index; BNT= Boston Naming Test; LDF = lower dose frequency; HDF = higher dose frequency; IMP = impairment-based treatment; A/P = activity/participation-based treatment; INT= integrated treatment

**Table 2.**

Quality Indicator Summative Scores for Included Studies

| Design | Test | N | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|---|---|---|---|
| Within Group | WAB | 53 | N/A | 2 | 17 | 21 | 32 | 28 | 0 |
| | CETI | 17 | N/A | 12 | 24 | 35 | 67 | 18 | 0 |
| | BNT | 36 | N/A | 6 | 11 | 28 | 33 | 22 | 0 |
| Between Group | WAB | 6 | 50 | 33 | 17 | 0 | 0 | 0 | 0 |
| | BNT | 5 | 0 | 20 | 80 | 0 | 0 | 0 | 0 |

*Note:* Value in cell represents percentage of studies with that summative score. Within group studies could not obtain a rating of 7 because intention to treat is not a relevant parameter for that study design. Higher scores = higher methodological quality.