**Melodic Intonation Therapy for aphasia: A multi-level meta-analysis**

Tudor Popescu1,2\*, Benjamin Stahl3-6, Brenton M. Wiernik7, Hannah Helm3, Michaela Zemanek1, Felix Haiduk1, Theresa Matzinger1,8, Roland Beisteiner2, Tecumseh W. Fitch1

1 University of Vienna, Department of Behavioural and Cognitive Biology, Vienna, Austria

2 Medical University of Vienna, Department of Neurology, Vienna, Austria

3 Medical School Berlin, Department of Psychology, Berlin, Germany

4 Charité Universitätsmedizin, Department of Neurology, Berlin, Germany

5 Max Planck Institute for Human Cognitive and Brain Sciences, Department of Neurophysics, Leipzig, Germany

6 Universitätsmedizin Greifswald, Department of Neurology, Greifswald, Germany

7 University of South Florida, Department of Psychology, Tampa, Florida, USA

8 University of Vienna, Department of English, Vienna, Austria

\*: tudor.popescu@univie.ac.at

# Abstract

INTRODUCTION

Melodic Intonation Therapy (MIT) is a prominent rehabilitation programme for individuals with post-stroke aphasia. Despite substantial progress in recent years, the overall efficacy and clinical relevance of MIT remain not fully understood: available studies are often constrained by lack of standardised outcomes, randomisation or control group. These limitations challenge the interpretation of the data.

AIMS

The present meta-analysis seeks to determine the efficacy and clinical relevance of MIT while considering quality of outcomes (psychometrically validated *versus* unvalidated measures), experimental design (presence *versus* absence of randomisation and control group), influence of spontaneous recovery (quantified as number of months post-stroke), MIT version applied (original *versus* modified protocol), and level of generalisation (testing performance on trained *versus* untrained items).

METHODS

An extensive literature search led to 606 studies being identified, 22 of which met the eligibility criteria. Multi-level mixed- and random-effects models were used to separately meta-analyse randomised controlled trial (RCT) and non-RCT data.

REULTS

Unvalidated outcomes appeared to attenuate the effect size of MIT by a factor of 0.29–0.43 across study designs when compared to validated outcomes. Additionally, MIT's effect size was 5.7 times larger for non-RCT data compared to RCT data. Effect size also decreased with number of months post-stroke, suggesting confounds related to spontaneous recovery, primarily within the first year post-stroke. In contrast, variation of the original MIT protocol did not systematically alter the benefit from treatment. Crucially, analyses demonstrate significantly improved language performance on trained and untrained items; the latter finding arose mainly from gains in repetition tasks, but not in other domains of verbal expression, including everyday communication ability.

CONCLUSIONS

Accounting for various methodological aspects, the current results confirm the promising role of MIT in improving language performance on trained items and in repetition tasks, while highlighting possible limitations in promoting everyday communication ability.

# Introduction

Stroke survivors often experience a profound loss of communication skills, among them a syndrome known as aphasia. This syndrome may manifest as severe difficulty in verbal expression, referred to as ‘non-fluent aphasia.’ In addition, stroke survivors frequently suffer from impaired speech-motor planning. Known as ‘apraxia of speech,’ this syndrome typically occurs in combination with aphasia. Although about a third of individuals with neurological communication disorders do not recover completely (Engelter et al., 2006), rehabilitation programmes can improve language performance even in the chronic stage of symptoms (Breitenstein et al., 2017).

Melodic Intonation Therapy (MIT) is a prominent rehabilitation programme originally developed for individuals with non-fluent aphasia (Albert et al., 1973). Drawing on the observation that individuals with neurological communication disorders are often able to sing entire pieces of text fluently (Gerstman, 1964; Mills, 1904; Yamadori et al., 1977), MIT uses melody, rhythm, vocal expression (in unison and alone), left-hand tapping, formulaic and non-formulaic verbal utterances, as well as other therapeutic elements, in a hierarchically structured protocol (Helm-Estabrooks et al., 1989). To date, randomised controlled trial (RCT) data have confirmed the efficacy of MIT on validated outcomes in the *late subacute* or *consolidation* stage of aphasia (i.e., up to 12 months after stroke; van der Meulen et al., 2014); but not in the *chronic* stage of aphasia (i.e. more than 6–12 months after stroke; Van Der Meulen et al., 2016).

From a methodological point of view, influences of spontaneous recovery are generally lower in the *chronic* stage of aphasia, as suggested by RCT data (Doppelbauer et al., 2021) and meta-analyses (RELEASE Collaborators, 2021). Therefore, it is important to consider stage of symptoms post-stroke. Moreover, speech-language therapy seeks to promote performance on untrained items. Consistent with this goal, the present work distinguishes progress on trained items—learning resulting from using the same set of utterances both during treatment and subsequent assessment—from the more desirable goal of attaining generalisation to untrained items, ideally in the context of everyday communication to ensure ecological validity (e.g., Blomert et al., 1994).

So far, to our knowledge there are several systematic reviews on MIT (e.g., van der Meulen et al., 2012; Zumbansen et al., 2014) and two meta-analyses (Brady et al., 2016; Zumbansen & Tremblay, 2018). These meta-analyses reflect a relatively limited amount of RCT data (Brady et al., 2016) or dichotomise post-treatment improvement in a way that prevents specific estimates of effect size (Zumbansen & Tremblay, 2018). Given the substantial burden of disease associated with aphasia, the present meta-analysis attempts to provide a deeper understanding about the potential and limitations of MIT. To achieve this goal, the current analyses synthesise available studies on MIT to address five research questions:

1. **Psychometric quality of outcomes.** Does the use of validated *versus* unvalidated outcomes systematically alter the resulting effect size of MIT?
2. **Experimental design.** Do RCT and non-RCT results in the context of MIT differ systematically in terms of effect size?
3. **Aphasia stage.** Do influences related to spontaneous recovery, quantified as number of months post-stroke-onset (MPO), affect the effect size of MIT?
4. **Variants of MIT protocol.** Do variations of the original MIT protocol alter the resulting effect size?
5. **Generalisability.** Apart from trained items, does MIT enhance performance on untrained items and, if so, does the resulting effect size demonstrate gains on measures of everyday communication ability?

# Methods

## Eligibility criteria

### Inclusion criteria

We defined the following basic inclusion criteria for studies to be considered for the present meta-analysis:

* empirical study that administered MIT to adult individuals (age 18 or over) with aphasia, with or without a control group;
* language-related outcomes in pre-post assessment;
* publication in peer-reviewed journal.

Despite the limited quality of evidence in non-RCT designs, we chose to also include case reports with individual patient data (IPD), analysing those studies separately and comparatively relative to RCTs.

### Exclusion criteria

After the removal of duplicate items (see *Literature search procedure* in the Supplementary Materials), the following exclusion criteria were applied to remaining studies:

* substantial variation from original MIT protocol (Albert et al., 1973). We accepted *minor* changes to the MIT protocol (and examined the effect of the categorical variable: original *versus* modified MIT), as long as the protocol met all of the following features:
  + - melody-based vocal expression;
    - some form of rhythmic pacing (e.g., left-hand tapping);
    - use of verbal utterances known from everyday communicative interaction;
* unvalidated outcome measures; no published or otherwise accessible validation study for the particular test battery. Exception: if a study included both trained and untrained items for an unvalidated measure, we included it to determine the degree of generalisation by comparing performance on trained and untrained items;
* other essential data not reported and / or not retrievable, even after contacting the authors (e.g., no sample size or standard error, insufficient information to compute an effect size).

### Included and excluded studies

For a list of the included studies, and also of the excluded (though initially included) studies, see the Supplementary Materials.

## Search strategy

The literature search strategy was devised and implemented by authors TP and MZ. This was designed to obtain high search sensitivity, using both free-text and subject headings in databases, which were not restricted by language or publication form (Higgins et al., 2021). The full literature search procedure, including search terms and databases used, and the PRISMA statement diagram, is reported in the Supplementary Materials.

## Study coding and double-coding

All studies were coded by the first author (TP). Two of the authors (FH, TM) re-coded all studies, verifying the cross-coder consistency. Agreement among the three coders occurred in a majority of cases, and any discrepancies found between coding sheets were solved by consensus. The ICCs (intraclass correlations) were >0.9 in the remaining cases, such as errors arisen from numerically estimating data reported in plot format only.

## Tests and outcome measures in primary studies

Tables in the Supplementary Materials show

* A hierarchical categorisation scheme showing how, for each of the target syndromes considered (aphasia and apraxia of speech), the different measures (subtests) from the batteries of validated tests that we considered contributed towards the relevant linguistic *Abilities*, and further towards the meta-analysed dependent variables (which we deemed *Domains*).
* All the tests reported in the primary studies considered, and the reason for excluding some of them.

## MIT variants

The included MIT variants were: SIPARI, Music therapy (sic!), Singing therapy, Speech–Music Therapy for Aphasia (SMTA), Modified Melodic Intonation Therapy (MMIT). The excluded variants (cf exclusion criteria) were: choir therapy, metrical pacing technique (MPT), music therapy combined with SLT, Modified Melodic Intonation Therapy (MMIT).

## Meta-analysis methods

### Computed outcome metric

To maximise comparability of effects across studies, we used change scores from pre-test to post-test as the outcome variable, expressed in *z*-scores. For group-level studies (the RCTs in the current analyses), we standardised *z*-scores using pooled pre-test standard deviation across control and treatment groups. For individual patient data studies (the case reports in the current analyses), we computed *z*-scores in one of three ways. For studies that reported results as *z*-scores (e.g., based on test norms), we used the *z*-scores directly. For studies that reported results as percentile scores (e.g., based on test norms), we converted these to *z*-scores using the quantiles of the standard Normal distribution. For other studies, we estimated *z*-scores using the following procedure. We first converted normalised raw scores to reflect the proportion of the maximum possible score (POMP; Cohen et al., 1999).[[1]](#footnote-1) Next, we estimated a three-level random-intercept model for the pre-test POMP scores, with individual test scores nested within patients nested within studies (see Figure 1). From these models, we used the population intercept as the estimated POMP score *mean*, and the patient-level random effects standard deviation as the estimated POMP score *SD* (τ). We then used this *mean* and *SD* to standardise the pre-test and post-test POMP scores.



Figure 1: Three-level model employed, with SDs (τ) at study-, patient- and measure-level.

### RCTs

All RCTs were reported at the group-level. We computed effect sizes as the pretest-posttest-control group Hedges’ *g*:  (Morris, 2008). We computed the variance for each *g* using the method of Morris (2008). We estimated multi-level mixed effects meta-regression models to account for effect size dependency, with random intercepts for each study. We first fit an overall meta-analysis combining all effect sizes. Second, we fit additional meta-regression models including potential moderator variables. For these meta-regression models, we included random slopes for the *Domain* moderator, nested within studies (Konstantopoulos, 2011). We used a homoscedastic compound symmetric structure for the random effects, estimating a single random effects variance and correlation for all abilities.[[2]](#footnote-2) We estimated the amount of heterogeneity (i.e., *τ*) using the restricted maximum-likelihood estimator (Viechtbauer, 2005). We computed confidence intervals for meta-regression coefficients and mean treatment effects using the Knapp and Hartung *t*-distribution method (Knapp & Hartung, 2003), and for the random effects components using profile likelihood. We estimated models using *R* (version 4.1.0; R Core Team, 2021) and the *metafor* package (version 3.-01 Viechtbauer, 2010).[[3]](#footnote-3)

### Case reports

All case reports reported results as individual-level data, so we analysed these studies using individual patient data (IPD) meta-analysis. We computed individual-level scores as the difference between pre-test and post-test *z*-scores (the mean difference in these scores is the pretest-posttest Hedges’ *g*, *gpp*). We then pooled data across studies using a three-level random-effects IPD meta-analysis, with individual scores again (see Figure 1) nested within patients nested within studies (Riley et al., 2010). Similar to the group-level RCT meta-analyses, we first fit an overall model including all data points with no moderators, then fit additional models including potential moderator variables as predictors. For these models, we included random intercepts for patients and studies.[[4]](#footnote-4) We estimated random effects components using REML and computed confidence intervals using profile likelihood. We estimated models using *R* (version 4.1.0; R Core Team, 2021) and the *lme4* package (version 1.1-27; Bates et al., 2015a).

### Moderator analyses

For the RCT meta-analyses, we fit a meta-regression model with the moderators (1) Domain (cf. section 2.4); (2) whether the study used validated tests as its outcome measures, or unvalidated ones (for unvalidated measures, we treated trained and untrained items as separate groups to avoid confounding measure validation and training effects); and (3) the Domain × Validated interaction. Next, we fit another model adding the additional moderators of (1) mean MPO across treatment and control groups; and (2) the difference in mean MPO between treatment and control groups.

For the case report meta-analyses, we initially fitted the same meta-regression model with three moderators as for RCTs. We then fit two additional models adding one moderator at a time to this baseline model. First, we fit a model adding individual-level MPO. Second, we fit a model adding whether a study used the original MIT protocol or a modified protocol.

# Results

Study-level standardised mean difference scores and meta-analytic mean differences by Domain are shown in Figure 2. Full meta-regression results tables are reported in the Supplementary Materials.

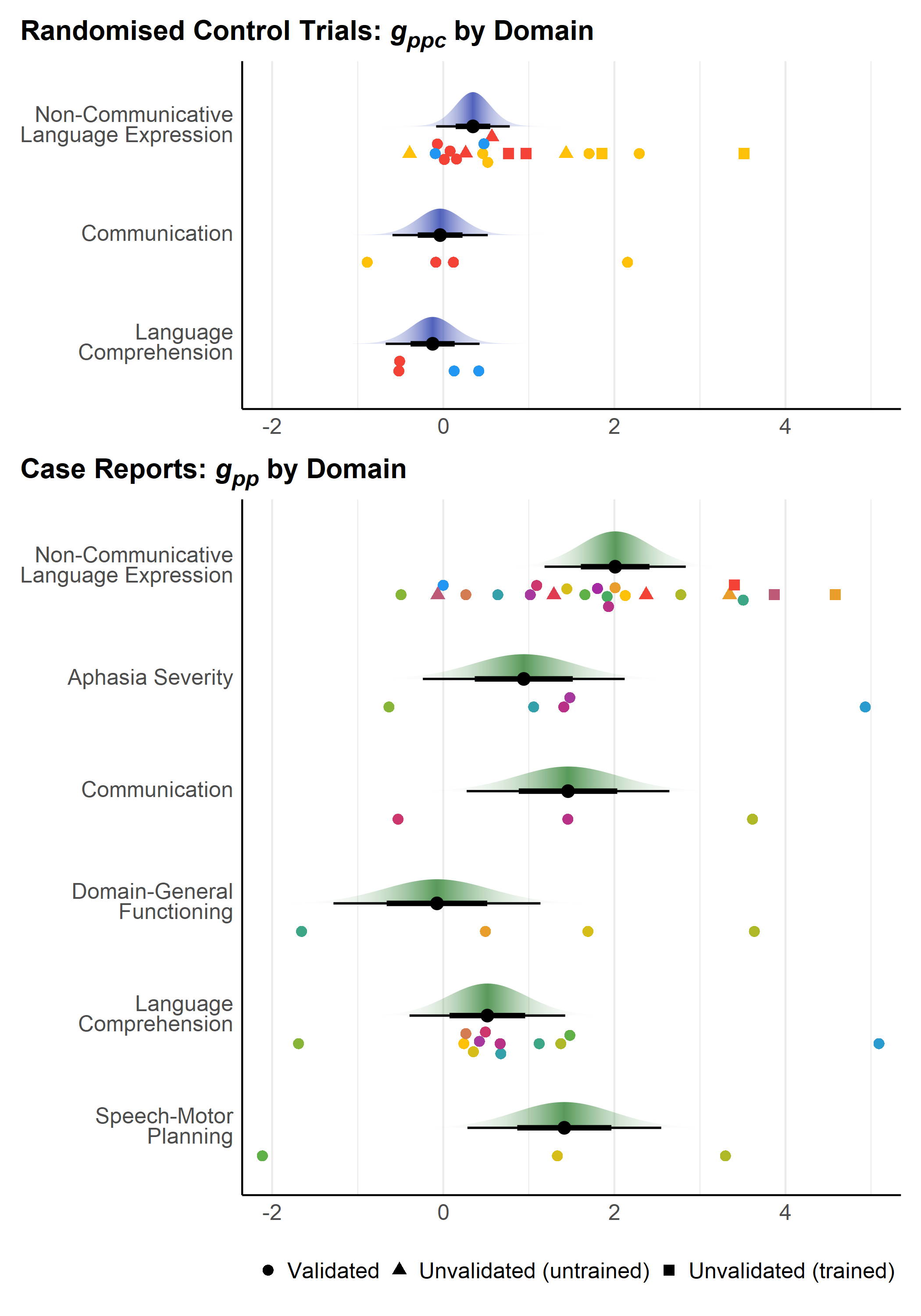


Figure 2: Results of meta-analyses. Points are study-level standardised mean pretest-posttest difference scores, either adjusted for a control group (*gppc*) or not (*gpp*). Points of different colours are drawn from different studies. Large points are mean *gpp(c)* for validated measures with 66% (thick bar) and 95% (thin bar) confidence intervals and *t*-distribution confidence densities. For case reports, one aphasia severity study with *gppc* = −4.88 is not displayed.

## RCT data

Overall, RCT data showed a small to moderate pretest-posttest effect of MIT on aphasia outcomes, after accounting for the control group (*g̅* = .31 [95% CI −.01, .63]). These results were primarily based on Non-Communicative Language Expression (repetition) tasks. Other abilities were less commonly assessed. In moderator analyses, effects appeared to be much weaker for Communication and Language Comprehension tasks than for Non-Communicative Language Expression, but confidence intervals for these differences were wide (see Figure 2). Effects were estimated to be somewhat heterogeneous across studies (random effects standard deviation, *τ =*.33 [95% CI .15, 1.01]).

Two studies included several unvalidated measures of Non-Communicative Language Expression. For these measures, treatment effects for untrained items were somewhat smaller than those for validated measures, though the confidence interval for this difference was fairly wide (∆*g̅* = −.15 [95% CI −.46, .15]). As expected, estimated treatment effects were much larger when patients were tested using trained items (∆*g̅* = .99 [95% CI .60, 1.39]; trained vs. untrained items contrast: 1.15 [95% CI .74, 1.56]). Smaller effect sizes for unvalidated measures may be attributable to poorer reliability compared to validated measures; measurement error tends to attenuate effect sizes (Ivanova & Hallowell, 2013; van Smeden et al., 2020; Wiernik & Dahlke, 2020).

When aphasia stage (MPO) was added to the RCT model, neither mean MPO across groups (∆*g̅* per month = −.008 [95% CI −.024, .008]) nor difference in mean MPO between MIT and control groups (∆*g̅* per month = −.004 [95% CI −.020, .011]) showed meaningful relationships with MIT treatment effects. Importantly, effect sizes for RCT analyses were drawn from only three studies, so these group-level MPO analyses have limited power to estimate the impact of MPO on MIT treatment effects.

## Case report data

Compared to RCT studies, case reports with no control group estimated much larger effects of MIT (*g̅* = 1.72 [95% CI 1.00, 2.42]). As with RCT studies, these results were primarily based on Non-Communicative Language Expression (repetition) tasks. Overall aphasia severity and language comprehension appeared to show somewhat smaller effects, but confidence intervals on these differences were very wide. Effects were estimated to be highly heterogeneous across studies (*τ* [between-studies] = 1.41 [95% CI .89, 2.05]),), to the degree that MIT was even estimated to be harmful in a small proportion of settings (for instance, the 95% normal-theory prediction interval for Non-Communicative Language Expression ranged −0.88 to +4.90; IntHout et al., 2016).

Four studies included several unvalidated measures of Non-Communicative Language Expression. As with RCT studies, treatment effects for untrained items on unvalidated measures appeared to be smaller than those for validated measures (with a wide confidence interval; ∆g̅ = −.47 [95% CI −2.40, 1.46]). Also similar to RCTs, apparent treatment effects were much larger for trained items (∆g̅ = 2.37 [95% CI .44, 4.31]; trained vs. untrained items contrast: 2.84 [95% CI 1.21, 4.48]).

When aphasia stage (MPO) was added to the case reports model, MPO showed a moderate negative relationship with treatment effects (∆*g̅* per month = −.02 [95% CI −.03, −.01]; estimated effect for 12 months, −.18 [95% CI −.30, −.07]; estimated effect for 24 months, −.37 [95% CI −.61, −.14]).

Compared to studies that used the original MIT protocol, studies that used a modified variant of the protocol appeared to show somewhat larger treatment effects, though the confidence interval on this difference was very wide (∆*g̅* = .56 [95% CI −.92, 2.03]).

# Discussion

The present meta-analysis aimed to investigate the efficacy and clinical relevance of MIT while accounting for crucial methodological aspects of primary studies, such as validated outcomes, use of control comparison, and randomised group allocation. Our results reveal that several poor methodological practices in this literature may introduce substantial biases into estimated treatment effects. For example, for RCT studies of non-communicative language expression, using unvalidated outcome measures (using untrained items) may attenuate observed MIT treatment effects by about 43% (*g̅unvalidated* = .20 vs. *g̅validated* = .35) when compared to validated outcome measures. Similarly, holding domain and validation status constant, we estimate the apparent effect size of MIT to be ≈ 5.7 times larger (*g̅case report* = 2.01 vs. *g̅RCT* = .35) for non-RCT data compared to RCT data (for validated Non-Communicative Language Expression measures).

## Research implications

These results indicate that rigorous study design (e.g., random assignment, control groups, high-quality measurement) can help reduce confounds and obtain more realistic estimates. In particular, this result reaffirms that (having and adjusting for) a control group is critical. Without this, most of the change observed in case reports, visible in inflated estimates of the efficacy of MIT, are inseparable from the phenomena of spontaneous recovery as well as, ultimately, regression to the mean, none of which are due to the treatment. Effect size also decreased with number of months post-stroke for IPD studies, indicating that progress in language performance reported in the late subacute or consolidation stage of aphasia may confound MIT effects with influences of spontaneous recovery, much of which is likely to already have occurred by that stage. Taken together, these results suggest that validated outcomes, randomised-controlled designs and inclusion of individuals with chronic aphasia are essential prerequisites to determine the efficacy in a reliable way.

## Clinical implications

Interestingly, variation of the original MIT protocol did not systematically alter the effect size, challenging the idea that modification of the treatment necessarily diminishes its outcome. However, few of the included studies employed an MIT variant, and their individual effects are heterogeneous. Our results can therefore express no certainty about the impact of deviations from the original MIT protocol, and instead highlight the need for more rigorous research investigating the effects of specific deviations.

Although MIT appears to benefit language performance on trained items, gains on untrained items emerged mostly from progress in repetition tasks. In contrast, the current results failed to demonstrate comparable progress in everyday communication ability on validated outcomes. These findings have implications in clinical practice.

## Limitations and future directions

As with any meta-analysis, the conclusiveness of the results strongly depends on the quality of the source material. As always, methodological shortcomings of primary studies emphasise the need for caution in interpreting the results. Nonetheless, the present meta-analysis considered various methodological aspects that tended to be neglected in previous work. In particular, our meta-analyses carefully determined the psychometric quality of each outcome, relative to recently defined standards in aphasia research (Wallace et al., 2019). In addition, the analyses accounted for quality of the research design, in terms of using control interventions and group randomisation to address unspecific influences, including bias due to placebo effects. Our results demonstrate that the overall efficacy of MIT in repetition tasks appears to persist, albeit to a smaller degree than previously reported.

Interestingly, variation of the original MIT protocol did not systematically alter the effect size, challenging the idea that modification of the treatment necessarily diminishes its outcome. This finding casts doubt on the notion that the original composition and hierarchical structure of MIT are indispensable for improving language performance. However, few of the included studies employed an MIT variant, and their individual effects are heterogeneous. Our results can therefore express no certainty about the impact of deviations from the original MIT protocol, and instead highlight the need for more rigorous research investigating the effects of specific deviations.

Using unvalidated outcomes, cross-sectional and longitudinal multiple-case studies have examined the role of different MIT elements: melody and rhythm (e.g., Kershenbaum et al., 2019), vocal expression in unison or alone (Racette et al., 2006), left-hand tapping (e.g., Zipse et al., 2014), and formulaicity of verbal utterances (e.g., Stahl et al., 2011). Possible methodological reasons for seemingly contradictory data, as well as conjectured mechanism of MIT, have been discussed (Stahl & Kotz, 2014). Obviously, the present results do not offer insight into any of these mechanisms. If indeed adherence to the original MIT protocol does not manifest in significantly elevated language performance, our results encourage future research to optimise the composition and structure of the treatment, to increase its efficacy in the rehabilitation of neurological communication disorders. For example, individuals with apraxia of speech may benefit from several elements of MIT, such as rhythmic pacing (Aichert et al., 2016) and language formulaicity (Stahl et al., 2020).

We here present the first meta-analysis on MIT that attempts to rule out various methodological caveats in interpreting the outcome of the treatment, such as lack of validated outcomes, control group or randomisation in previous studies. Accounting for each of these issues in a rigorous way, the results of our meta-analyses confirm the promising role of MIT in improving language performance on trained items and in repetition tasks, while highlighting possible limitations in promoting everyday communication ability. We hope that the current work will be helpful for clinicians, patients and families to make informed decisions about their treatment options to support recovery from post-stroke aphasia.

# Acknowledgements

We gratefully acknowledge:

* Joost Hurkmans for providing unpublished work.
* Sarah Wallace for providing test-related information.
* Luisa Krein and Emily Braun for information regarding test scores.
* Ajay Halai and Yina Quique Buitrago for very useful comments.

# Author contributions (CRediT author statement)

Conceptualization: TP, BS, WTF

Methodology: TP, BMW

Software: TP, BMW

Validation: TP

Formal analysis: TP, BMW

Investigation: TP

Resources: TP, RB, WTF

Data Curation: TP, BMW, FH, TM

Writing - Original Draft: TP

Writing - Review & Editing: TP, BS, BMW, FH, TM, RB, WTF

Visualization: TP

Supervision: TP, BS

Project administration: TP

Funding acquisition: RB, WTF

additional ones not included in the CRediT author statement

literature search and curation: TP, HH, MZ

# References

Aichert, I., Späth, M., & Ziegler, W. (2016). The role of metrical information in apraxia of speech. Perceptual and acoustic analyses of word stress. *Neuropsychologia*, *82*, 171–178. https://doi.org/10.1016/j.neuropsychologia.2016.01.009

Albert, M. L., Sparks, R. W., & Helm, N. A. (1973). Melodic Intonation Therapy for Aphasia. *Archives of Neurology*, *29*(2), 130–131. https://doi.org/10.1001/archneur.1973.00490260074018

Bates, D. (2006, May 19). *[R] lmer, p-values and all that*. https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015a). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015b). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Blomert, L., Kean, M. L., Koster, C., & Schokker, J. (1994). Amsterdam—Nijmegen everyday language test: Construction, reliability and validity. *Aphasiology*, *8*(4), 381–407. https://doi.org/10.1080/02687039408248666

Brady, M. C., Kelly, H., Godwin, J., & Enderby, P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews*, *5*. https://doi.org/10.1002/14651858.CD000425.pub3

Breitenstein, C., Grewe, T., Flöel, A., Ziegler, W., Springer, L., Martus, P., Huber, W., Willmes, K., Ringelstein, E. B., Haeusler, K. G., Abel, S., Glindemann, R., Domahs, F., Regenbrecht, F., Schlenck, K.-J., Thomas, M., Obrig, H., de Langen, E., Rocker, R., … Bamborschke, S. (2017). Intensive speech and language therapy in patients with chronic aphasia after stroke: A randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. *The Lancet*, *389*(10078), 1528–1538. https://doi.org/10.1016/S0140-6736(17)30067-3

Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*, *34*(3), 315–346. https://doi.org/10.1207/S15327906MBR3403\_2

Doppelbauer, L., Mohr, B., Dreyer, F. R., Stahl, B., Büscher, V., & Pulvermüller, F. (2021). Long-Term Stability of Short-Term Intensive Language-Action Therapy in Chronic Aphasia: A 1-2 year Follow-Up Study. *Neurorehabilitation and Neural Repair*, 15459683211029236. https://doi.org/10.1177/15459683211029235

Engelter, S. T., Gostynski, M., Papa, S., Frei, M., Born, C., Ajdacic-Gross, V., Gutzwiller, F., & Lyrer, P. A. (2006). Epidemiology of aphasia attributable to first ischemic stroke: Incidence, severity, fluency, etiology, and thrombolysis. *Stroke*, *37*(6), 1379–1384. https://doi.org/10.1161/01.STR.0000221815.64093.8c

Gerstman, H. L. (1964). A case of aphasia. *The Journal of Speech and Hearing Disorders*, *29*, 89–91. https://doi.org/10.1044/jshd.2901.89

Helm-Estabrooks, N., Morgan, A. R., & Nicholas, M. (1989). *Melodic intonation therapy*. Pro-Ed.

Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (Eds.). (2021). *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane. www.training.cochrane.org/handbook

Hurkmans, J., Bruijn, M. de, Boonstra, A. M., Jonkers, R., Bastiaanse, R., Arendzen, H., & Reinders-Messelink, H. A. (2012). Music in the treatment of neurological language and speech disorders: A systematic review. *Aphasiology*, *26*(1), 1–19. https://doi.org/10.1080/02687038.2011.602514

IntHout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, *6*(7), e010247. https://doi.org/10.1136/bmjopen-2015-010247

Ivanova, M. V., & Hallowell, B. (2013). A tutorial on aphasia test development in any language: Key substantive and psychometric considerations. *Aphasiology*, *27*(8), 891–920. https://doi.org/10.1080/02687038.2013.805728

Kershenbaum, A., Nicholas, M. L., Hunsaker, E., & Zipse, L. (2019). Speak along without the song: What promotes fluency in people with aphasia? *Aphasiology*, *33*(4), 405–428. https://doi.org/10.1080/02687038.2017.1413487

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, *22*(17), 2693–2710. https://doi.org/10.1002/sim.1482

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis: Three-level meta-analysis. *Research Synthesis Methods*, *2*(1), 61–76. https://doi.org/10.1002/jrsm.35

Mills, C. K. (1904). Treatment of aphasia by training. *Journal of the American Medical Association*, *43*(26), 1940–1949.

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364–386. https://doi.org/10.1177/1094428106291059

R Core Team. (2021). *R: A language and environment for statistical computing* (4.1.0) [Computer software]. R Foundation for Statistical Computing. http://www.r-project.org/

Racette, A., Bard, C., & Peretz, I. (2006). Making non-fluent aphasics speak: Sing along! *Brain*, *129*(10), 2571–2584. https://doi.org/10.1093/brain/awl250

RELEASE Collaborators. (2021). Predictors of Poststroke Aphasia Recovery: A Systematic Review-Informed Individual Participant Data Meta-Analysis. *Stroke*, *52*(5), 1778–1787. https://doi.org/10.1161/STROKEAHA.120.031162

Revicki, D., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, *61*(2), 102–109. https://doi.org/10.1016/j.jclinepi.2007.03.012

Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct, and reporting. *BMJ*, *340*, c221. https://doi.org/10.1136/bmj.c221

Stahl, B., Gawron, B., Regenbrecht, F., Flöel, A., & Kotz, S. A. (2020). Formulaic Language Resources May Help Overcome Difficulties in Speech-Motor Planning after Stroke. *PloS One*, *15*(6), e0233608. https://doi.org/10.1371/journal.pone.0233608

Stahl, B., & Kotz, S. A. (2014). Facing the music: Three issues in current research on singing and aphasia. *Frontiers in Psychology*, *5*, 1033. https://doi.org/10.3389/fpsyg.2014.01033

Stahl, B., Kotz, S. A., Henseler, I., Turner, R., & Geyer, S. (2011). Rhythm in disguise: Why singing may not hold the key to recovery from aphasia. *Brain*, *134*(10), 3083–3093. https://doi.org/10.1093/brain/awr240

Stahl, B., Mohr, B., Dreyer, F. R., Lucchese, G., & Pulvermüller, F. (2017). Communicative-Pragmatic Assessment Is Sensitive and Time-Effective in Measuring the Outcome of Aphasia Therapy. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00223

van der Meulen, I., van de Sandt-Koenderman, M. E., & Ribbers, G. M. (2012). Melodic Intonation Therapy: Present Controversies and Future Opportunities. *Archives of Physical Medicine and Rehabilitation*, *93*(1, Supplement), S46–S52. https://doi.org/10.1016/j.apmr.2011.05.029

Van Der Meulen, I., Van De Sandt-Koenderman, M. W. M. E., Heijenbrok, M. H., Visch-Brink, E., & Ribbers, G. M. (2016). Melodic Intonation Therapy in Chronic Aphasia: Evidence from a Pilot Randomized Controlled Trial. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00533

van der Meulen, I., van de Sandt-Koenderman, W. Mieke. E., Heijenbrok-Kal, M. H., Visch-Brink, E. G., & Ribbers, G. M. (2014). The Efficacy and Timing of Melodic Intonation Therapy in Subacute Aphasia. *Neurorehabilitation and Neural Repair*, *28*(6), 536–544. https://doi.org/10.1177/1545968313517753

Van Lancker Sidtis, D. (2021). *Foundations of Familiar Language: Formulaic Expressions, Lexical Bundles, and Collocations at Work and Play* (1st edition). Wiley-Blackwell.

van Smeden, M., Lash, T. L., & Groenwold, R. H. H. (2020). Reflection on modern methods: Five myths about measurement error in epidemiological research. *International Journal of Epidemiology*, *49*(1), 338–347. https://doi.org/10.1093/ije/dyz251

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261–293. https://doi.org/10.3102/10769986030003261

Viechtbauer, W. (2010). Conducting meta-analyses in *R* with the *metafor* package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., Babbitt, E., Bose, A., Brady, M., Cherney, L. R., Copland, D., Cruice, M., Enderby, P., Hersh, D., Howe, T., Kelly, H., Kiran, S., Laska, A.-C., Marshall, J., … Webster, J. (2019). A core outcome set for aphasia treatment research: The ROMA consensus statement. *International Journal of Stroke*, *14*(2), 180–185. https://doi.org/10.1177/1747493018806200

Wiernik, B. M., & Dahlke, J. A. (2020). Obtaining unbiased results in meta-analysis: The importance of correcting for statistical artifacts. *Advances in Methods and Practices in Psychological Science*, *3*(1), 94–123. https://doi.org/10.1177/2515245919885611

Yamadori, A., Osumi, Y., Masuhara, S., & Okubo, M. (1977). Preservation of singing in Broca’s aphasia. *Journal of Neurology, Neurosurgery, and Psychiatry*, *40*(3), 221–224. https://doi.org/10.1136/jnnp.40.3.221

Zipse, L., Worek, A., Guarino, A. J., & Shattuck-Hufnagel, S. (2014). Tapped out: Do people with aphasia have rhythm processing deficits? *Journal of Speech, Language, and Hearing Research: JSLHR*, *57*(6), 2234–2245. https://doi.org/10.1044/2014\_JSLHR-L-13-0309

Zumbansen, A., Peretz, I., & Hébert, S. (2014). Melodic Intonation Therapy: Back to Basics for Future Research. *Frontiers in Neurology*, *5*. https://doi.org/10.3389/fneur.2014.00007

Zumbansen, A., & Tremblay, P. (2018). Music-based interventions for aphasia could act through a motor-speech mechanism: A systematic review and case–control analysis of published individual participant data. *Aphasiology*, *33*(4), 466–497. https://doi.org/10.1080/02687038.2018.1506089

# Supplementary Materials [[to be made into separate doc?]]

## 

## Literature search procedure

The systematic literature search in literature databases yielded 606 hits; through searching the trials registers 7 additional trials were found. These 613 items have been subjected to a duplicate check for identical publications found through the different search tools combined. They were checked with the “duplicate detection” feature of the reference management software Zotero, with the following procedure: When all of the duplicate items had the same publication form and all of the duplicates had an abstract, we kept the first one in the list (by order of importing into Zotero); if not, we kept the first item of the duplicates having an abstract. With different publication entries from the same study we followed the preference hierarchy: journal article > book chapter > conference proceedings. After this process of eliminating all supplicates 143 items remained. Following this first step we read and checked all abstracts of the remaining articles against the exclusion criteria. Within this step 44 articles were excluded leaving 99 articles that checked all eligibility criteria. As the final step the full texts of all remaining articles were reviewed. During two rounds, 78 articles were excluded leaving 21 final articles that fit all our inclusion criteria and could therefore be included in this meta analysis.

See also the Prisma flow chart in Appendix N, which summarises the study counts above.

### Electronic searches

We searched the following databases: *Cochrane Library* (last searched 25.02.2021), *CINAHL EBSCOhost* (27.02.2021), *PsycINFO OVID* (1806 to August 2020), *Web of Science* (25.02.2021), *PubMed* (26.02.2021), *Scopus* (26.02.2021), *Medline* *OVID* (1946 to February 2021) *PSYNDEX OVID* (1977 to February 2021), *Music Periodicals Database* (26.02.2021), and *ProQuest Dissertations & Theses Global* (27.02.2021).

We also searched trials registers, including *International Clinical Trials Registry Platform* (ICTRP,<https://www.isrctn.com/>; 11.03.2021), *National Research Register* (UK),<http://www.nihr.ac.uk/>; 11.03.2021), *Clinical Trials.gov* ([www.clinicaltrials.gov](http://www.clinicaltrials.gov); 11.03.2021), *Netherlands Trials Register*[www.trialregister.nl](http://www.trialregister.nl); 11.03.2021), and the *German clinical trials Register*<https://www.drks.de/drks_web/>; 11.03.2021)

Additionally, we performed searches in Google Scholar (11.03.2021) and in the grey literature database OpenGrey.eu (<http://www.opengrey.eu/>; 11.03.2021). Messages soliciting any unpublished data were additionally sent to:

* aphasia associations
  + National Aphasia Association, NAA, https://www.aphasia.org
  + Australian Aphasia Association (AAA), https://aphasia.org.au
  + Fédération Nationale des Aphasiques de France (FNAF), http://aphasie.fr/
* music therapy associations
  + American Music Therapy Association (AMTA), https://www.musictherapy.org/
  + British Association for Music Therapy (BAMT), https://www.bamt.org
* mailing lists
  + AUDITORY
  + Musicology-all
* authors of the included studies

Finally, to ensure no studies were omitted we consulted the list of studies in published systematic reviews and meta-analyses concerning MIT (Brady et al., 2016; Hurkmans et al., 2012; van der Meulen et al., 2012; Zumbansen et al., 2014; Zumbansen & Tremblay, 2018).

Search histories and counts of the number of search hits and items remaining during subsequent filtering steps, are provided in the Supplementary Materials.

Since no filters relating to methods used or publication type were applied to our searches, we manually separated and kept the empirical studies from the overall results.

### Search terms

*CINAHL:*

((MM "Aphasia+") OR "aphasia" OR (MH "Aphasia, Broca") OR (MH "Aphasia, Transcortical Sensory") OR (MH "Aphasia, Wernicke")) AND (((MH singing or singing ) AND ("speech therapy" or (MH "Speech Therapy+")) OR "melodic intonation therapy")))

ClinicalTrials.gov

"melodic intonation therapy"

*Cochrane Library*

"melodic intonation therapy" AND aphasia

("speech therapy“ in Ti Abstr Key OR MeSH descriptor [speech therapy] explode all trees) AND singing\* in Ti Abstr Key

*Deutsches Register klinischer Studien (DRKS)/German clinical trials register*<https://www.drks.de/drks_web/>

melodic intonation therapy

melodische intonationstherapie

*Google Scholar*

„melodic intonation therapy“

„melodische intonationstherapie“

*ICTRP (International Clinical Trials Registry Platform)*

"melodic intonation therapy"

*Medline OVID*

("melodic intonation therapy".ab. or "melodic intonation therapy".ti. or "melodic intonation therapy".id. or ((singing.ab. or singing.id. or singing.ti. or singing/) AND (speech therapy/ or "speech therapy".ab. or "speech therapy".id. or "speech therapy".ti.))) AND (aphasia.ab. or aphasia.id. or aphasia.ti. or exp aphasia)

*Music Periodicals Database*

"melodic intonation therapy" and aphasia

*National Research Register (UK):* [*http://www.nihr.ac.uk/*](http://www.nihr.ac.uk/)

"melodic intonation therapy"

*Netherlands Trials Register www.trialregister.nl*

Melodic intonation

*OpenGrey.eu http://www.opengrey.eu/*

„melodic intonation therapy“

*ProQuest Dissertations & Theses Global*

„melodic intonation therapy“

*PsycINFO OVID*

("melodic intonation therapy".ab. or "melodic intonation therapy".ti. or "melodic intonation therapy".id. or ((singing.ab. or singing.id. or singing.ti. or singing/) AND (speech therapy/ or "speech therapy".ab. or "speech therapy".id. or "speech therapy".ti.))) AND (aphasia.ab. or aphasia.id. or aphasia.ti. or exp aphasia)

*PSYNDEX OVID*

("melodic intonation therapy".ab. or "melodic intonation therapy".ti. or "melodic intonation therapy".id. or ((singing.ab. or singing.id. or singing.ti. or singing/) AND (speech therapy/ or "speech therapy".ab. or "speech therapy".id. or "speech therapy".ti.))) AND (aphasia.ab. or aphasia.id. or aphasia.ti. or exp aphasia)

*PubMed*

(„melodic intonation therapy“ OR ((singing [MeSH Terms] OR singing) AND ("speech therapy"[MeSH Terms] OR "speech therapy")) AND (aphasia[MeSH Terms] OR aphasia)

*Scopus*

((( TITLE-ABS-KEY ( singing ) AND (TITLE-ABS-KEY ( "speech therapy" OR "language therapy" ))) OR ( TITLE-ABS-KEY ( "melodic intonation therapy" ))) AND ( TITLE-ABS-KEY ( aphasia))

*Web of Science*

TOPIC: (("melodic intonation therapy") OR TOPIC: ("singing" AND ("speech therapy" OR "language therapy")) AND TOPIC: aphasia

### PRISMA statement diagram

We used the PRISMA statement as found at http://prisma-statement.org/prismastatement/Checklist.aspx.

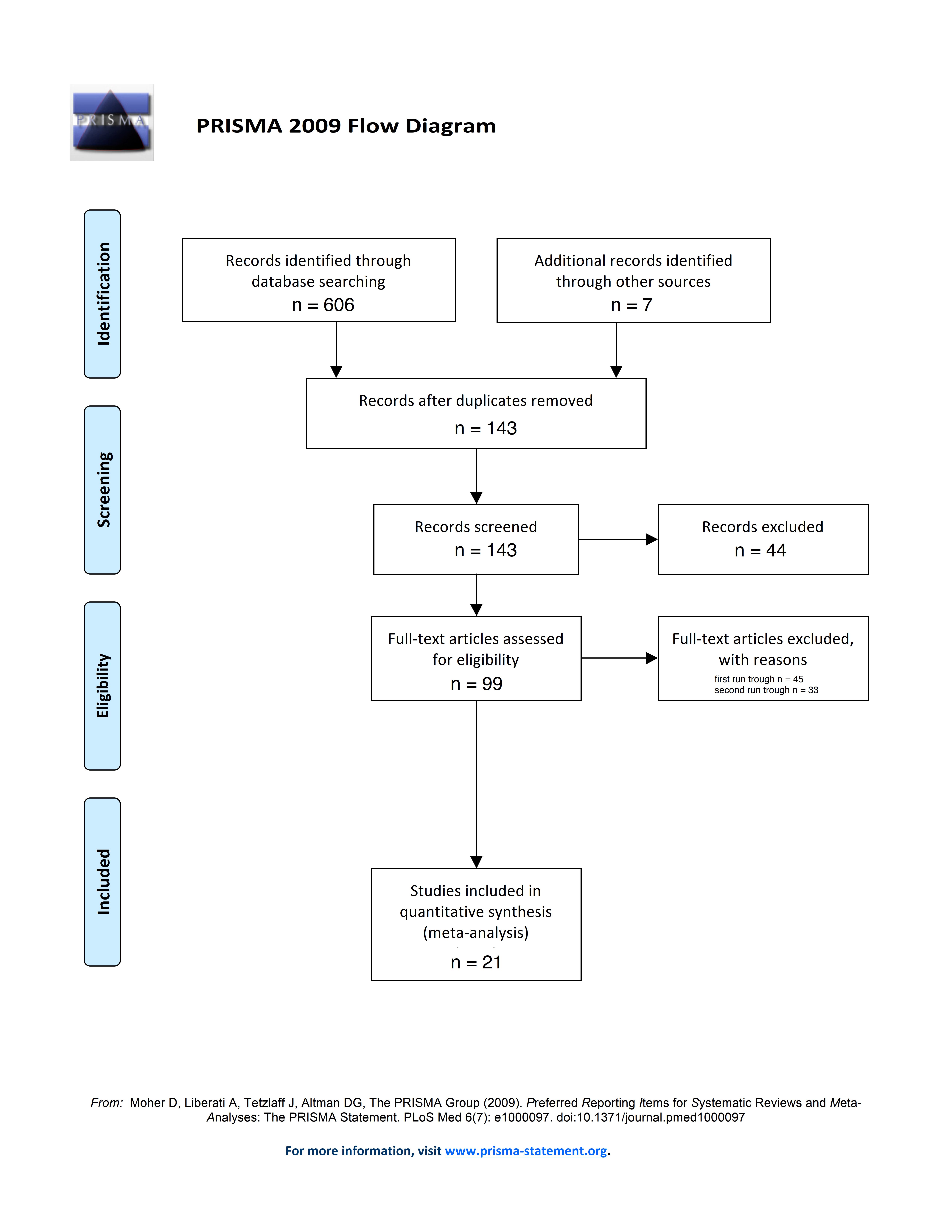


Figure 3: PRISMA statement and flow diagram, cf http://prisma-statement.org/prismastatement/Checklist.aspx.

## Lists of primary studies considered

### Included studies

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Study type** | **1st auth** | **Year** | **Title** | **N (MIT group)** | **N (ctrl group)** | **IPD** | **MIT variant** |
| case series | Akanuma | 2016 | Singing can improve speech function in aphasics associated with intact right basal ganglia and preserve right temporal glucose metabolism: Implications for singing therapy indication | 10 | - | 1 | singing therapy |
| case series | Belin | 1996 | Recovery from nonfluent aphasia after melodic intonation therapy: A PET study | 7 | - | 1 | TMR |
| single case | Bitan | 2018 | Changes in Resting-State Connectivity following Melody-Based Therapy in a Patient with Aphasia | 1 | - | 1 | MMIT |
| uncontrolled before and after study | Cortese | 2015 | Rehabilitation of aphasia: application of melodic-rhythmic therapy to Italian language | 6 | - | 1 | MRT |
| case series | Haro-Martínez | 2017 | Adaptation of melodic intonation therapy to Spanish: a feasibility pilot study | 4 | - | 1 | MIT |
| RCT | Haro-Martínez | 2019 | Melodic intonation therapy in post-stroke nonfluent aphasia: a randomized pilot trial | 20 | 20 | 0 | MIT |
| single case | Hatayama | 2021 | Music intonation therapy is effective for speech output in a patient with non‐fluent aphasia in a chronic stage | 1 | - | 1 | MIT |
| single case | Homan | 2015 | A Combination of Therapeutic Techniques: Severe Broca’s Aphasia | 1 | - | 1 | MMIT |
| case series | Hurkmans | 2015 | The effectiveness of Speech–Music Therapy for Aphasia (SMTA) in five speakers with Apraxia of Speech and aphasia | 5 | - | 1 | SMTA |
| case series | Jungblut | 2014 | Paving the Way for Speech: Voice-Training-Induced Plasticity in Chronic Aphasia and Apraxia of Speech—Three Single Cases | 3 | - | 1 | SIPARI |
| case series | Naeser | 1985 | CT Scan Lesion Localization and Response to Melodic Intonation Therapy with Nonfluent Aphasia Cases | 8 | - | 1 | MIT |
| single case | Primassin | 2014 | Melodische Intonationstherapie bei einer aphasischen Patientin in der (Post-) Akutphase | 1 | - | 1 | MIT |
| single case | Slavin | 2018 | A Case Study Using a Multimodal Approach to Melodic Intonation Therapy | 1 | - | 1 | SMTA |
| controlled before and after study | Stahl | 2013 | How to engage the right brain hemisphere in aphasics without even singing: evidence for two paths of speech recovery | 5 | - | 1 | singing therapy |
| single case | Tabei | 2016 | Improved Neural Processing Efficiency in a Chronic Aphasia Patient Following Melodic Intonation Therapy: A Neuropsychological and Functional MRI Study | 1 | - | 1 | MIT |
| single case | van de Sandt-Koenderman | 2010 | A Case Study of Melodic Intonation Therapy (MIT) in the Subacute Stage of Aphasia: Early Re-re activation of Left Hemisphere Structures | 1 | - | 1 | MIT |
| uncontrolled before and after study | van de Sandt-Koenderman | 2018 | Language lateralisation after Melodic Intonation Therapy: an fMRI study in subacute and chronic aphasia | 9 | - | 1 | MIT |
| case series | van der Meulen | 2012 | Melodic Intonation Therapy: Present Controversies and Future Opportunities | 2 | - | 1 | MIT |
| RCT | van der Meulen | 2014 | The Efficacy and Timing of Melodic Intonation Therapy in Subacute Aphasia | 23 | 25 | 0 | MIT |
| RCT | van der Meulen | 2016 | Melodic Intonation Therapy in Chronic Aphasia: Evidence from a Pilot Randomized Controlled Trial | 16 | 17 | 0 | MIT |
| single case | Wilson | 2006 | Preserved singing in aphasia: a case study of the eﬃcacy of the Melodic Intonation Therapy | 1 | - | 1 | palliative MIT (pMIT) |
| case series | Zumbansen | 2014 | The Combination of Rhythm and Pitch Can Account for the Beneficial Effect of Melodic Intonation Therapy on Connected Speech Improvements in Broca’s Aphasia | 3 | - | 1 | MIT |

### Excluded studies

Table 1: List of initially included, later excluded studies.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Study** | | | | **Reason for exclusion** | | | | | | |
| **Type** | **First author** | **Year** | **Title** | **Substantial variation from original MIT protocol** | **Non-validated tests** | **No contrast of trained vs untrained items** | **No pre&post data** | **Other essential data not reported (&unretrievable after contacting authors)** | **Untrustworthy source, e.g. non-peer-reviewed journal** | **Patient(s) 17 or younger** |
| case series | Al-Janabi | 2014 | Augmenting melodic intonation therapy with non-invasive brain stimulation to treat impaired left-hemisphere function: two case studies |  | 1 | 1 | 1 |  |  |  |
| case series | Baker | 2000 | Modifying the Melodic Intonation Therapy Program for Adults With Severe Non-fluent Aphasia |  | 1 |  |  |  |  |  |
| case series | Breier | 2010 | Changes in maps of language activity activation following melodic intonation therapy using magnetoencephalography: Two case studies |  | 1 | 1 |  |  |  |  |
| case series | Kim | 2008 | Protocol Evaluation for Effective Music Therapy for Persons with Nonfluent Aphasia |  | 1 | 1 | 1 |  |  |  |
| case series | Mauszycki | 2016 | Melodic intonation therapy applied to the production of questions in aphasia |  | 1 | 1 |  |  |  |  |
| case series | Schlaug | 2009 | Evidence for plasticity in white-matter tracts of patients with chronic Broca’s aphasia undergoing intense intonation-based speech therapy. |  | 1 |  |  |  |  |  |
| case series | Schlaug | 2008 | From Singing to Speaking: Why Singing May Lead to Recovery of Expressive Language Function in Patients with Broca's Aphasia |  | 1 | 1 |  |  |  |  |
| case series | Sparks | 1974 | Aphasia rehabilitation resulting from melodic intonation therapy |  |  |  |  | 1 |  |  |
| case series | Wambaugh | 2012 | Acquired Apraxia of Speech: The Effects of Repeated Practice and Rate/Rhythm Control Treatments on Sound Production Accuracy | 1 |  |  |  |  |  |  |
| controlled before and after study | Wan | 2014 | Intensive therapy induces contralateral white matter changes in chronic stroke patients with Broca’s aphasia |  | 1 |  |  |  |  |  |
| controlled before and after study | Osisanya | 2012 | Effectiveness of melodic intonation therapy in the management of communication difficulty of pupils with non-fluent aphasia in the classroom setting |  |  |  |  |  | 1 |  |
| Cross-over trial | Brendel | 2008 | Effectiveness of metrical pacing in the treatment of apraxia of speech | 1 |  |  |  |  |  |  |
| RCT | Conklyn | 2012 | The Effects of Modified Melodic Intonation Therapy on Nonfluent Aphasia: A Pilot Study |  | 1 | 1 |  |  |  |  |
| RCT | Raglio | 2016 | Improvement of spontaneous language in stroke patients with chronic aphasia treated with music therapy: a randomized controlled trial | 1 |  |  |  |  |  |  |
| RCT | Zumbansen | 2017 | Effect of choir activity in the rehabilitation of aphasia: a blind, randomised, controlled pilot study | 1 |  |  |  |  |  |  |
| RCT | Vines | 2011 | Non invasive brain stimulation enhances the effects of melodic intonation therapy | 1 |  | 1 |  |  |  |  |
| single case | Hough | 2010 | Melodic Intonation Therapy and aphasia: another variation on a theme |  | 1 |  |  |  |  |  |
| single case | Jungblut | 2009 | Long-term recovery from chronic global aphasia: A case report |  |  |  | 1 |  |  |  |
| single case | Lagasse | 2012 | Evaluation of Melodic Intonation Therapy for Developmental Apraxia of Speech |  |  |  |  | 1 |  |  |
| single case | Zipse | 2012 | When right is all that is left: plasticity of right-hemisphere tracts in a young aphasic patient |  |  |  |  |  |  | 1 |
| single case | Fountura | 2014 | Efficacy of the Adapted Melodic Intonation Therapy: a case study of a Broca`s Aphasia Patient |  |  |  |  |  | 1 |  |
| single-case study | Goldfarb | 1979 | Espousing melodic intonation therapy in aphasia rehabilitation: a case study. |  | 1 | 1 | 1 |  |  |  |
| single-case study | Morrow-Odom | 2013 | Effectiveness of melodic intonation therapy in a case of aphasia following right hemisphere stroke |  | 1 |  |  |  |  |  |
| single-case study | Wambaugh | 2000 | Eﬀects of rate and rhythm control treatment on consonant production accuracy in apraxia of speech | 1 |  |  |  |  |  |  |
| uncontrolled before and after study | Bonakdarpour | 2003 | Melodic intonation therapy in Persian aphasic patients |  | 1 | ? |  |  |  |  |
| uncontrolled before and after study | Lim | 2013 | The Therapeutic Effect of Neurologic Music Therapy and Speech Language Therapy in Post-Stroke Aphasic Patients |  |  |  | 1 | 1 |  |  |
| uncontrolled before and after study | Hurkmans | 2016 | The treatment of apraxia of speech: Speech and music therapy, an innovative joint effort | 1 |  |  |  |  |  |  |
|  | Keith | 1975 | Singing as therapy for apraxia of speech and aphasia: Report of a case | 1 |  |  |  |  |  |  |
|  | Krauss | 1982 | Melodic intonation therapy with language delayed apraxic children |  |  |  |  |  |  | 1 |
|  | Marshall | 1976 | Melodic Intonation Therapy: Variations on a Theme |  |  | 1 |  |  |  |  |
|  | Mauszycki | 2008 | The effects of rate control treatment on consonant production accuracy in mild apraxia of speech | 1 |  |  |  |  |  |  |
|  | Springer | 1993 | Training in the use of wh-questions and prepositions in dialogues: A comparison of two different approaches in aphasia therapy |  | 1 |  |  | 1 |  |  |
|  | Tonkovich | 1977 | The Effects of Stress and Melodic Intonation on Apraxia of Speech |  |  |  |  |  |  |  |

## Tests and outcome measures in primary studies

Table 2: List of all tests considered.

|  |  |  |
| --- | --- | --- |
| **Test abbrev** | **Test full name** | **Validation study exists** |
| - | Farsi aphasia test | No |
| - | Apraxia Battery for Adults | No |
| AABT | Aachener Aphasie Bedside Test | Yes |
| AAT | Aachen Aphasia Test | Yes |
| ADP | Aphasia Diagnostic Proﬁles | No |
| ANELT | Amsterdam-Nijmegen Everyday Language Test | Yes |
| BDAE | Boston Diagnostic Aphasia Examination | Yes |
| BNT | Boston Naming Test | Yes |
| CAL | Communicative Activity Log | No |
| CIU | Sabadel CIUs/min | No |
| DIAS | Diagnostic Instrument for Apraxia of Speech (AoS) | Yes\* |
| HWL | Hierarchical Word List | Yes\* |
| MT86 | Montréal–Toulouse Aphasia Battery | No |
| PALPA | Psycholinguistic assessments of language processing in aphasia | Yes |
| PICA | Porch Index Of Communicative Ability | Yes |
| PPTT | Pyramids and Palm Trees Test | Yes |
| SLTA | Standard Language Test of Aphasia | Yes |
| WAB | Western Aphasia Battery | Yes |

Table 3: Categorisation scheme showing how, for each target syndrome, Subtests → Tests → Abilities → Domains.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Target syndrome | DOMAIN | Ability | Test battery | Subtest |
| Aphasia | APHASIA SEVERITY | Overall language performance | AAT | aphasia severity |
| Aphasia | APHASIA SEVERITY | Overall language performance | BDAE | aphasia severity |
| Aphasia | APHASIA SEVERITY | Overall language performance | PALPA | (no particular subtests) |
| Aphasia | APHASIA SEVERITY | Overall language performance | PPTT | (no particular subtests) |
| Aphasia | APHASIA SEVERITY | Overall language performance | WAB | aphasia quotient (AQ) |
| Aphasia | COMMUNICATION | Everyday communication | ANELT | comprehensibility |
| Aphasia | COMMUNICATION | Everyday communication | ANELT | intelligibility |
| Aphasia | COMMUNICATION | Everyday communication | ANELT | verbal communication |
| Aphasia | DOMAIN-GENERAL FUNCTION | Cognitive-Executive Skills | AABT | BLIKO = Aufforderungen zu Blick- und Kopfbewegungen |
| Aphasia | DOMAIN-GENERAL FUNCTION | Cognitive-Executive Skills | AABT | IDENT = Identifizieren von Objekten |
| Aphasia | DOMAIN-GENERAL FUNCTION | Cognitive-Executive Skills | AABT | MUMO = Aufforderungen zu Mundbewegungen |
| Aphasia | DOMAIN-GENERAL FUNCTION | Cognitive-Executive Skills | AAT | token |
| Aphasia | LANGUAGE COMPREHENSION | Auditory Comprehension | AAT | auditory comprehension |
| Aphasia | LANGUAGE COMPREHENSION | Auditory Comprehension | BDAE | auditory commands |
| Aphasia | LANGUAGE COMPREHENSION | Auditory Comprehension | BDAE | auditory comprehension |
| Aphasia | LANGUAGE COMPREHENSION | Auditory Comprehension | BDAE | complex auditory material |
| Aphasia | LANGUAGE COMPREHENSION | Auditory Comprehension | WAB | auditory comprehension |
| Aphasia | LANGUAGE COMPREHENSION | Written Comprehension | AAT | written comprehension |
| Aphasia | LANGUAGE COMPREHENSION | Written Comprehension | AAT | written language |
| Aphasia | LANGUAGE COMPREHENSION | Written Comprehension | BDAE | reading comprehension |
| Aphasia | LANGUAGE COMPREHENSION | Written Comprehension | WAB | reading |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Articulatory Agility | BDAE | articulatory agility |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Grammatical Form | BDAE | grammatical form |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Naming | AABT | BENENN = Benennen von Objekten |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Naming | AAT | naming |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Naming | BDAE | naming |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Naming | BDAE | naming (confrontation ~) |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Naming | BDAE | naming (responsive ~) |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Naming | BNT | naming |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Naming | SLTA | picture (manga) description test |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Naming | WAB | naming |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Phrase Length | BDAE | phrase length |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Repetition | (trained) | repetition |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Repetition | (untrained) | repetition |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Repetition | AABT | SIREI = Singen, Reihen- und Floskelsprechen |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Repetition | AAT | repetition |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Repetition | BDAE | repetition |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Repetition | WAB | repetition |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Spontaneous Speech | AAT | automatic language |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Spontaneous Speech | AAT | communication |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Spontaneous Speech | AAT | phonetic language |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Spontaneous Speech | AAT | prosody |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Spontaneous Speech | AAT | semantic language |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Spontaneous Speech | AAT | syntactic language |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Spontaneous Speech | WAB | spontaneous speech |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Syllable production | (trained) | correct syllables across test-phrases |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Syllable production | (trained) | correct syllables per test-phrase |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Syllable production | (untrained) | correct syllables across test-phrases |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Syllable production | (untrained) | correct syllables per test-phrase |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Verbal Expression | BDAE | verbal expression |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Word production | (trained) | proportion of words correct |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Word production | (untrained) | proportion of words correct |
| Aphasia | NON-COMMUNICATIVE LANGUAGE EXPRESSION | Writing | WAB | writing |
| Apraxia of Speech | SPEECH-MOTOR PLANNING | Speech-Motor Planning | DIAS | articulation of phonemes |
| Apraxia of Speech | SPEECH-MOTOR PLANNING | Speech-Motor Planning | DIAS | articulation of words |
| Apraxia of Speech | SPEECH-MOTOR PLANNING | Speech-Motor Planning | DIAS | diadockokinesis (DDK) |
| Apraxia of Speech | SPEECH-MOTOR PLANNING | Speech-Motor Planning | HWL | number of assessable items |
| Apraxia of Speech | SPEECH-MOTOR PLANNING | Speech-Motor Planning | HWL | phonemic structure |
| Apraxia of Speech | SPEECH-MOTOR PLANNING | Speech-Motor Planning | HWL | phonetic structure |
| Apraxia of Speech | SPEECH-MOTOR PLANNING | Speech-Motor Planning | HWL | speech fluency |

## Results tables

### Overall RCT meta-analyses

Table 4: Overall RCT meta-analyses.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Term | Estimate | *SE* | Statistic | *df* | *p* | 95% conf. int. |
| *g̅* | 0.31 | 0.16 | 2.00 | 25 | 0.057 | [−0.01, 0.63] |
| τ | 0.25 |  | 212.08 | 25 | < 0.001 | [ 0.10, 1.11] |

*Note.* *g̅* = mean pretest-posttest difference (*gppc*; accounting for control group); τ = estimated random effects standard deviation across studies; Statistic = *t* value for *g̅* and *QE* value for τ; confidence intervals computed using *t* distributions for *g̅* and profile likelihood for τ and ρ.

### RCT meta-analyses of broad ability categories

Table 5: RCT meta-analyses of broad ability categories.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Term | Estimate | *SE* | Statistic | *df* | *p* | 95% conf. int. |
| *g̅* (Non-communicative Language Expression) | 0.35 | 0.21 | 1.68 | 21 | 0.108 | [-0.08, 0.78] |
| *g̅* (Communication) | -0.04 | 0.27 | -0.14 | 21 | 0.893 | [-0.59, 0.52] |
| *g̅* (Language Comprehension) | -0.12 | 0.26 | -0.47 | 21 | 0.643 | [-0.67, 0.42] |
| ∆*g̅* (unvalidated measure with untrained items) | -0.15 | 0.15 | -1.06 | 21 | 0.300 | [-0.46, 0.15] |
| ∆*g̅* (unvalidated measure with trained items) | 0.99 | 0.19 | 5.24 | 21 | < .001 | [ 0.60, 1.39] |
| τ | 0.33 |  | 158.71 | 21 | < .001 | [ 0.15, 1.01] |
| ρ | -0.05 |  |  | 21 |  | [-0.52, 0.93] |

*Note.* *g̅* = mean pretest-posttest difference (*gppc*; accounting for control group); τ = estimated random effects standard deviation across studies; ρ = estimated correlation among *g* treatment effects between measures of different ability domains across studies; Statistic = *t* value for *g̅* and *QE* value for τ; confidence intervals computed using *t* distributions for *g̅* and profile likelihood for τ.

### Overall IPD meta-analyses

Table 6: Overall IPD meta-analyses.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Term | Estimate | *SE* | *t* | 95% conf. int. |
| *g̅* | 1.43 | 0.30 | 4.84 | [0.82, 2.02] |
| τ | 1.00 |  |  | [0.59, 1.55] |
| σ (person) | 0.93 |  |  | [0.73, 1.20] |
| σ (measure) | 1.45 |  |  | [1.38, 1.52] |

*Note.* *g̅* = mean pretest-posttest difference (*gpc*; not accounting for any control group); τ = estimated random effects standard deviation across studies; σ (person) = estimated random effects standard deviation across persons (within study); σ (measure) = estimated random effects standard deviation across measures (within person); confidence intervals computed using profile likelihood; *p* values omitted as the appropriate denominator degrees of freedom for linear mixed effects models is ill-defined (Bates, 2006; Bates et al., 2015b, pp. 34–35); inference should be based on the profile likelihood confidence intervals.

### IPD meta-analyses of broad ability categories

Table 7: IPD meta-analyses of broad ability categories.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Term | Estimate | *SE* | *t* | 95% conf. int. |
| *g̅* (Non-communicative Language Expression) | 1.45 | 0.35 | 4.12 | [ 0.74, 2.14] |
| *g̅* (Aphasia Severity) | 0.51 | 0.44 | 1.16 | [-0.37, 1.38] |
| *g̅* (Communication) | 1.59 | 0.39 | 4.08 | [ 0.81, 2.35] |
| *g̅* (Domain-General Function) | 1.19 | 0.40 | 2.94 | [ 0.38, 1.97] |
| *g̅* (Language Comprehension) | 0.26 | 0.37 | 0.70 | [-0.50, 0.99] |
| *g̅* (Speech-Motor Planning) | 1.41 | 0.38 | 3.68 | [ 0.64, 2.15] |
| ∆*g̅* (unvalidated measure with untrained items) | 0.31 | 0.75 | 0.41 | [-1.16, 1.80] |
| ∆*g̅* (unvalidated measure with trained items) | 1.86 | 0.75 | 2.50 | [ 0.40, 3.35] |
| τ | 1.15 |  |  | [ 0.70, 1.69] |
| σ (person) | 0.93 |  |  | [ 0.73, 1.19] |
| σ (measure) | 1.39 |  |  | [ 1.32, 1.46] |

*Note.* *g̅* = mean pretest-posttest difference (*gpc*; not accounting for any control group); ∆*g̅* = estimated difference in *g̅* between validated and unvalidated measures; note that only Non-communicative Language Expression included unvalidated measures; τ = estimated random effects standard deviation across studies; σ (person) = estimated random effects standard deviation across persons (within study); σ (measure) = estimated random effects standard deviation across measures (within person); confidence intervals computed using profile likelihood; *p* values omitted as the appropriate denominator degrees of freedom for linear mixed effects models is ill-defined (Bates, 2006; Bates et al., 2015b, pp. 34–35); inference should be based on the profile likelihood confidence intervals.

### IPD meta-analyses with aphasia stage (MPO) as a moderator

Table 8: IPD meta-analyses with aphasia stage (months post-onset, MPO) as a moderator.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Term | Estimate | *SE* | *t* | 95% conf. int. |
| *g̅* (Non-communicative Language Expression) | 1.58 | 0.21 | 7.35 | [ 1.16, 1.99] |
| *g̅* (Aphasia Severity) | 0.88 | 0.42 | 2.07 | [ 0.05, 1.70] |
| *g̅* (Communication) | 2.08 | 0.26 | 8.07 | [ 1.57, 2.58] |
| *g̅* (Domain-General Function) | 2.07 | 0.29 | 7.13 | [ 1.51, 2.63] |
| *g̅* (Language Comprehension) | 0.59 | 0.26 | 2.25 | [ 0.08, 1.10] |
| *g̅* (Speech-Motor Planning) | 1.86 | 0.25 | 7.51 | [ 1.38, 2.34] |
| ∆*g̅* (unvalidated measure with untrained items) | 0.32 | 0.47 | 0.68 | [-0.60, 1.24] |
| ∆*g̅* (unvalidated measure with trained items) | 1.87 | 0.47 | 3.98 | [ 0.96, 2.79] |
| ∆*g̅* (per month post-onset) | -0.02 | 3.74e-03 | -4.19 | [-0.02, -0.01] |
| τ | 0.00 |  |  | [ 0.00, 0.51] |
| σ (person) | 0.98 |  |  | [ 0.75, 1.24] |
| σ (measure) | 1.30 |  |  | [ 1.23, 1.37] |

*Note.* *g̅* = mean pretest-posttest difference (*gpc*; not accounting for any control group); ∆*g̅* = estimated difference in *g̅*; note that only Non-communicative Language Expression included unvalidated measures; τ = estimated random effects standard deviation across studies; σ (person) = estimated random effects standard deviation across persons (within study); σ (measure) = estimated random effects standard deviation across measures (within person); confidence intervals computed using profile likelihood; *p* values omitted as the appropriate denominator degrees of freedom for linear mixed effects models is ill-defined (Bates, 2006; Bates et al., 2015b, pp. 34–35); inference should be based on the profile likelihood confidence intervals.

### IPD meta-analyses with MIT protocol as a moderator

Table 9: IPD meta-analyses with MIT protocol as a moderator.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Term | Estimate | *SE* | *t* | 95% conf. int. |
| *g̅* (Non-communicative Language Expression) | 1.35 | 0.82 | 1.64 | [-0.17, 2.86] |
| *g̅* (Aphasia Severity) | 0.21 | 0.92 | 0.22 | [-1.48, 1.90] |
| *g̅* (Communication) | 1.57 | 0.84 | 1.86 | [ 0.02, 3.13] |
| *g̅* (Domain-General Function) | 1.14 | 0.85 | 1.34 | [-0.43, 2.71] |
| *g̅* (Language Comprehension) | 0.05 | 0.84 | 0.05 | [-1.50, 1.59] |
| *g̅* (Speech-Motor Planning) | 1.34 | 0.84 | 1.60 | [-0.21, 2.89] |
| ∆*g̅* (unvalidated measure with untrained items) | -0.55 | 1.11 | -0.49 | [-2.59, 1.49] |
| ∆*g̅* (unvalidated measure with trained items) | 1.04 | 1.11 | 0.94 | [-1.00, 3.08] |
| ∆*g̅* (modified MIT protocol) | 0.79 | 0.96 | 0.82 | [-0.99, 2.57] |
| τ | 1.38 |  |  | [ 0.70, 2.01] |
| σ (person) | 0.98 |  |  | [ 0.75, 1.30] |
| σ (measure) | 1.41 |  |  | [ 1.34, 1.48] |

*Note.* *g̅* = mean pretest-posttest difference (*gpc*; not accounting for any control group); ∆*g̅* = estimated difference in *g̅*; note that only Non-communicative Language Expression included unvalidated measures; τ = estimated random effects standard deviation across studies; σ (person) = estimated random effects standard deviation across persons (within study); σ (measure) = estimated random effects standard deviation across measures (within person); confidence intervals computed using profile likelihood; *p* values omitted as the appropriate denominator degrees of freedom for linear mixed effects models is ill-defined (Bates, 2006; Bates et al., 2015b, pp. 34–35); inference should be based on the profile likelihood confidence intervals.

## Supplementary References

# [[To do for this manuscript]]

* check that reporting conforms to the guidelines in the
  + PRISMA statement (Moher, 2009)
  + cochrane handbook for systematic reviews of interventions, https://training.cochrane.org/handbook/current

1. For a small number of studies, it was not possible to determine the maximum or minimum possible scores. For these studies, we computed POMP scores using the maximum and minimum *observed* scores in the sample. Results did not change meaningfully if we excluded these studies from results. [↑](#footnote-ref-1)
2. For comparison, we also estimated models with unequal random effects variances across dependent variables. This did not improve model fit based on AICc comparison or likelihood ratio tests. [↑](#footnote-ref-2)
3. As only three RCT studies were identified, it was not possible to apply methods to detect publication-bias or other small-sample effects (e.g., tests of funnel plot asymmetry). [↑](#footnote-ref-3)
4. Models with random slopes for the *Domain* variable did not converge, likely due to the limited co-occurrence of specific pairs of those within any one study. [↑](#footnote-ref-4)