**Supplementary Materials**

The quality assessment revealed observable trends across the included studies. The majority of the studies failed to adequately report or control for potential bias within group allocation. Blinding efforts were largely insufficient for both participants and outcome assessors. The PEDro Scale additionally revealed that the majority of studies were unable to obtain data from more than 85% of their recruited cohort. However, it is worth noting that this is not uncommon in EEG studies of error processing, which require a certain number of errors for valid analysis and epochs of EEG data to be uncontaminated by artifacts, often resulting in the exclusion of a higher number of participants than in other studies. Additionally, the strengths of the studies were that the majority implemented and reported evidence-based selection criteria well. Whilst all 15 studies recorded a high risk of bias in the CCRBT (Fig. 12), it is worth noting that the set of domains in the tool is smaller, when compared with the PEDro Scale (Fig. 13) and the SQAC (Fig. 14). The PEDro Scale and SQAC included more criteria related to recruitment, study design and the reporting of results, which the studies scored strongly in overall.

According to the summary of PEDro measurement properties (Physiotherapy Evidence Database, 2021), five of the studies were classified as “good” (score 6 - 8), eight were “fair” (score 4 - 5) and two were “poor” (≤ 3). As the SQAC guidelines did not provide classifications of overall scores and studies were not excluded from data analysis based upon their score, the results were simply observed. Whilst the overall scores provided a clear indication of efforts to control for bias, they are limited by allowing cross-sectional designs to present higher scores. This is due to the exclusion of “N/A” judgements from the overall calculation. The excluded criteria (randomisation and blinding procedures) would strongly influence the quality of a study. As such, randomised controlled trials (RCTs) scored lower on average, even with the exclusion of Cragun et al. (2020) as a presumed outlier (RCTs: *M* = 0.83, cross-sectional: *M* = 0.91). The opposite result was observed from the PEDro scale, with RCTs scoring higher on quality (RCTs: *M* = 5.75, cross-sectional: *M* = 4). Note that Cragun et al. (2020) provided unpublished data for inclusion in the meta-analysis, so while we have included their study in our quality assessment for the sake of completeness, their total score was low, but influenced by missing information rather than necessarily being a poorly controlled study.

Whilst the quality assessment tools generally produced similar outcomes for individual sources of bias, the overall conclusions were still able to differ based on overall assessment calculations. This highlights the importance of utilising various tools to strengthen the validity of the quality assessment process.

Table

Description automatically generated

**Fig. 12** The results of the study quality assessment using the Cochrane Collaboration’s Tool for Assessing Risk of Bias (CCRBT; Higgins et al., 2011)

Note:

D1: Sequence generation.

D2: Allocation concealment.

D3: Blinding of participants and personnel.

D4: Blinding of outcome assessment.

D5: Incomplete outcome data.

D6: Selective reporting.

D7: Other sources of bias.

A picture containing chart

Description automatically generated

Fig. 13 The results of the study quality assessment using the PEDro Scale (1999)

Note:

D1: Eligibility Criteria were specified.

D2: Subjects were randomly allocated to groups.

D3: Allocation was concealed.

D4: Groups were similar at baseline regarding the most important indicators.

D5: There was blinding of all subjects.

D6: There was blinding of all therapists who administered the therapy.

D7: There was blinding of all assessors who measured at least one key outcome.

D8: Measures of at least one key outcome were obtained from more than 85% of the subjects initially allocated to groups.

D9: All subjects for whom outcome measures were available received the treatment or control condition as allocated or, where this was not the case, data for at least one key outcome was analysed by “intention to treat”.

D10: The results of between-group statistical comparisons are reported for at least one key outcome.

D11: The study provides both point measures and measures of variability for at least one key outcome.



Fig. 14 The results of the study quality assessment using the Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields (SQAC; Kmet et al., 2004)

Note:

D1: Is the objective of the study sufficiently described?

D2: Study design evident and appropriate?

D3: Method of subject/comparison group selection or source of information/input variables described and appropriate?

D4: Subject (and comparison group if applicable) characteristics sufficiently described?

D5: If interventional and random allocation was possible, was it described?

D6: If interventional and blinding of investigators was possible, was it reported?

D7: If interventional and blinding of subjects was possible, was it reported?

D8: Outcome and (if applicable) exposure measure(s) well defined and robust to measure/misclassification bias? Means of assessment reported?

D9: Sample size appropriate?

D10: Analytic methods described/justified and appropriate?

D11: Some estimate of variance is reported for the main results?

D12: Controlled for confounding?

D13: Result reported in sufficient detail?

D14: Conclusion supported by results?