UPDATE

Consensus on the reporting and experimental design of clinical and cognitive-behavioural neurofeedback studies (CRED-nf checklist)


These authors contributed equally to this work. All other authors are listed in reverse alphabetical order.

Neurofeedback has begun to attract the attention and scrutiny of the scientific and medical mainstream. Here, neurofeedback researchers present a consensus-derived checklist that aims to improve the reporting and experimental design standards in the field.

1 Departments of Neuroscience and Psychiatry, University of Geneva; Campus Biotech, Geneva, Switzerland
2 Department of Clinical Neuropsychology, University of Groningen, Groningen, The Netherlands
3 Department of Biomedical Sciences of Cells & Systems, University Medical Center Groningen, Groningen, The Netherlands
4 Laureate Institute for Brain Research, Tulsa, Oklahoma, USA
5 University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
6 Institute of Psychology, University of Graz, Graz, Austria
7 Massachusetts Institute of Technology, Cambridge, MA, USA

© The Author(s) (2020). Published by Oxford University Press on behalf of the Guarantors of Brain.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Introduction

After a protracted history, neurofeedback has begun to attract the attention and scrutiny of the scientific and medical mainstream (Kamiya, 2011; Linden, 2014; Sitaram et al., 2017). A debate now centres on the extent to which neurofeedback alters brain function and behaviour, and the mechanisms through which neurofeedback operates (e.g. neurofeedback-specific versus non-specific). A series of correspondences in *Lancet Psychiatry* (Micoulaud-Franchi and Fovet, 2016; Thibault and Raz, 2016a, b; Pigott et al., 2017; Schönemberg et al., 2017a, b) and *Brain* (Fovet et al., 2017; Schabus, 2017, 2018; Schabus et al., 2017; Thibault et al., 2017, 2018; Witte et al., 2018) discuss the theoretical arguments and empirical data backing the involvement of these two mechanisms.

The apparent controversy that the correspondence letters present stems from a well-known phenomenon in neuropsychology: that multiple components can drive the benefits of a treatment (Enriquez-Geppert et al., 2013; Campbell and Stanley, 2015). We depict this hypothesized multi-component model for the context of neurofeedback in Fig. 1. We divide the mechanisms driving experimental outcomes into five bins: neurofeedback-specific (related to training a target neurophysiological variable), neurofeedback non-specific (dependent on the neurofeedback context, but independent from the act of controlling a particular brain signal), general non-specific (including the common benefits of cognitive training as well as psychosocial influences, such as placebo responding), repetition related (e.g. test–retest improvement), and cognitive. 

Abbreviations: CRED-nf = Consensus on the reporting and experimental design of clinical and cognitive-behavioural neurofeedback studies; MCID = minimal clinically important difference
and natural (e.g. spontaneous remission, cognitive development) (Micolaud-Franchi and Fovet, 2018).

Although a framework based on these terms and concepts is only beginning to concretize in the neurofeedback literature, most scientists involved in neurofeedback agree on their general usage and interpretation. The greater points of contention centre on (i) whether previous experiments provide sufficient evidence to identify specific factors as a key driver of neurofeedback outcomes; and (ii) how to best design an experiment to clearly dissociate the various mechanisms driving neurofeedback outcomes. If neurofeedback outcomes occur independently of the information provided by the neural feedback signal (i.e. come from non-specific mechanisms), then neurofeedback does not rely on the main criteria that set it apart from other interventions, such as cognitive training and meditation. An ideal demonstration of neurofeedback-specific effects would include evidence of online (i.e. intra-session) and offline (i.e. inter-session or post-treatment) changes in targeted brain activity, as well as a control group or condition to rule out non-specific effects (e.g. sensory stimulation, placebo). Individual neurofeedback studies, however, contain varying proportions of each of these criteria and have led to a diversity of opinions regarding the specificity of mechanisms involved in neurofeedback. The present checklist provides the structure to develop a more comprehensive and rigorous evidence base.

Evidence for putatively causal, neurofeedback-specific mechanisms relies on our knowledge of the physiological basis of neural activity and its relevance to cognition (for a review of neurofeedback mechanisms, see Ros et al., 2014; Sitaram et al., 2017). For example, the association between neural activity and cognition in animals (Cao et al., 2016; Babapoor-Farrokhran et al., 2017) suggests that self-regulation of brain circuits can alter behaviour and cognition. A number of neurofeedback experiments in animals (Sterman et al., 1970; Schafer and Moore, 2011), and humans (Watanabe et al., 2017; Young et al., 2017b) further support this view. Evidence suggesting that mechanisms other than neurofeedback-specific factors account for the effects of neurofeedback come from a number of recent studies and reviews that find comparable benefits between participants who receive veritable neurofeedback from their own brain and those who observe a sham-neurofeedback signal unrelated to their neural activity of interest (Schabus et al., 2017; Schönenberg et al., 2017a; Thibault and Raz, 2017).

To advance the field of neurofeedback, scientists can benefit from designing future studies with the methodological rigour capable of disentangling the various mechanisms driving the effects of neurofeedback. As authors of the correspondence, alongside other researchers active in the field, we propose a standardized checklist outlining best practices in the experimental design and reporting of neurofeedback studies. We believe that widespread adoption of this checklist will help advance our scientific understanding of how neurofeedback affects brain function and behaviour.

**Objectives of the checklist**

This checklist is intended to encourage robust experimental design and clear reporting for clinical and cognitive-behavioural neurofeedback experiments (for a methodological review see Ros et al., 2014; Enriquez-Geppert et al., 2017). Because all neurofeedback aims to train brain activity, these guidelines generalize across EEG, magnetoencephalography (MEG), functional MRI, functional near infrared spectroscopy (fNIRS), and other neurofeedback modalities. The checklist focuses mainly on aspects unique to the neurofeedback context (as general standards for each imaging modality already exist; Gross et al., 2013; Nichols et al., 2017; Pernet et al., 2018). It serves as a complement, rather than alternative, to the Consolidated Standards of Reporting Trials (CONSORT) guidelines (Schulz et al., 2010) (http://www.consort-statement.org/checklists). When submitting neurofeedback results for publication, we encourage researchers to include the checklist (Fig. 2), ideally using the application available at www.rtfin.org/CREDnf. Alternatively, the checklist can be downloaded from the Supplementary material and the final column can be filled with the relevant text from your manuscript, or the page number identifying where in the manuscript each item is addressed. This checklist does not aim to inhibit the exploration of novel directions in neurofeedback research. On the contrary, it advocates robust designs and clear reporting to promote informed research decisions that can effectively build upon previous work. These guidelines are a first iteration. As neurofeedback research progresses, we invite the community to provide comments for improving this checklist (see rtf.in/CREDnf for a link to the commenting platform). We hope these guidelines will help disentangle the relative contribution of the mechanisms outlined in Fig. 1.

**Description of checklist items**

Below, we include a short description of each checklist item followed by examples from published neurofeedback articles.

**Pre-experiment**

**Item 1a. Preregister experimental protocol and planned analyses**

This item is essential for clinical and replication studies, and is encouraged for others.

Preregister, for example, on a platform such as www.osf.io, as a randomized controlled trial (RCT) on ClinicalTrials.gov or the European Union Clinical Trials Register (EUCTR), or by submitting a registered report (see www.cos.io/rr for information concerning registered reports). Clearly label primary and secondary outcome variables. Indicate the number, frequency, and duration...
of neurofeedback sessions. In the publication, report which analyses were preregistered, which were exploratory, and disclose any potential deviations from the preregistered protocol.

Examples:


(ii) See Holtmann et al. (2014) for a pre-published protocol of the study by Strehl et al. (2017) with trial registry number ISRCTN 76187185.

Item 1b. Justify sample size

This item is essential.

Describe the sampling plan and how it was determined. Ideally, justify the sample size with a power analysis based on the smallest effect size of interest [e.g. minimal clinically important differences (MCIDs), see Item 6a] or another method (e.g. Bayesian sequential sampling). Otherwise, label the experiment as a pilot, proof-of-concept, or feasibility study. If the preregistered sample size is not met, state so. Whereas smallest effect sizes of interest may be derived from previous literature, we do not recommend selecting a sample size based solely on an ‘expected’ effect size derived from previous published results. Because of publication bias, which remains common across research fields, this practice can leave experiments underpowered (Albers and Lakens, 2018; Algermissen and Mehler, 2018).

Examples:

(i) ‘Estimates of a clinically relevant effect size were derived from the Göttingen pilot-study using the same primary outcome measures [18]. It is expected that in the neurofeedback group the mean FBB-ADHS score at Post-Test 2 is 1.20 and in the control group 1.50 with a common standard deviation of 0.55. The expected outcome requires a sample size of 72 subjects per group (α = 0.05, two sample t-test, two-sided) to achieve a power of 90%.’ (Holtmann et al., 2014).

(ii) ‘Owing to feasibility and proof of principle, we intend following a Bayesian sampling strategy with a minimum of N=5 patients and continue recruiting either until the Bayes factor for both hypotheses (A and B) is conclusive - i.e. either for the alternative with BF10 > 10 (indicating strong evidence for a positive effect) or for the null with BF01 > 10 (indicating strong evidence for a null effect) - or until the end of the data collection period (September 30, 2017) is reached.’ (Mehler et al., 2017).

Control groups

Item 2a. Employ control group(s) or control condition(s)

This item is essential.

Use a control group (between subjects) or control condition (within subjects). This could include a placebo-control (e.g. sham-neurofeedback, neurofeedback from a largely unrelated brain signal, or inversing the neurofeedback reward contingency) or another active non-neurofeedback control (e.g. a similar type of computerized cognitive training, biofeedback, or medication). See Sorger et al. (2019) for an in-depth review of control groups in neurofeedback research. Consider the potential for, and report any adverse effects in both the experimental and control groups.

Examples:

(i) ‘Four separate healthy subject control groups were trained and tested using similar or identical procedures but in the absence of valid rACC rtfMRI information … Group III (n = 8) received identical training to the control group, but using rtfMRI information derived from a different brain region in posterior cingulate cortex that is not believed to be involved in pain processing to examine spatial and physiological specificity. Group IV (n = 4) received identical training to the experimental group, but, unknown to the subjects, the rtfMRI displays that they viewed corresponded to activation from a previously tested experimental subject’s rACC,
rather than their own rACC brain activation.’ (deCharms et al., 2005).

(ii) ‘As a semi-active control condition EMG feedback of coordination in the supraspinatus muscles was chosen. Participants were instructed either to contract or to relax the left relative to the right supraspinatus muscle. This protocol was chosen to induce differential EMG control corresponding to the “polarities” comparable to the NF condition, without requiring simple relaxation or tension. This allowed us to use the same device and the same representation of the feedback signal on the screen. We did not choose a standard EMG feedback protocol because the control condition should be as unspecific as possible but include the possibility to learn self-regulation, i.e. the unspecific variable of any biofeedback treatment.’ (Strehl et al., 2017).

**Item 2b. When leveraging experimental designs where a double-blind is possible, use a double-blind**

This item is essential.

For example, in experiments with a placebo-neurofeedback control group or within participant control conditions.

Example:

‘To blind staff to treatment condition, The SmartBox interface devices were independently preprogrammed by an off-site consultant who had no interaction with participants or data (analogous to prepackaged randomized medication).’ (Arnold et al., 2013).

Comment: Currently, few neurofeedback software packages are designed for blinding the treatment staff.

**Item 2c. Blind those who rate the outcomes, and when possible, the statisticians involved**

For this item, see Dutilh et al. (2019); this item is encouraged.

Indicate which individuals were blinded and how blinding was achieved.

Example:

‘The Behavioral Observation of Students in Schools [BOSS] … is a systematic interval recording observation system for coding classroom behavior and reports on engagement … and off-task behaviors … Data output from observations are objective quantitative assessments, which can help reduce observer bias … The BOSS was completed … for all study participants by trained RA [research assistants] who were unaware of the participants’ randomization conditions. The participants were unaware that they were being observed.’ (Steiner et al., 2014).

**Item 2d. Examine to what extent participants and experimenters remain blinded**

This item is encouraged.

For an overview on reporting whether blinding was successful, see Kolahi et al. (2009).

Example:

‘The CSQ [consumer satisfaction questionnaire], administered at Treatments 24 and 40, also included questions to examine blindness to treatment assignment … Of 34 participants at Treatment 40, 35% of children and 29% of parents said that they did not know which treatment they had been assigned to and declined to guess. Only 32% of children and 24% of parents guessed correctly, with 32% and 47%, respectively, guessing incorrectly.’ (Arnold et al., 2013).

**Item 2e. In clinical efficacy studies, employ a standard-of-care intervention group as a benchmark for improvement**

This item is encouraged.

This design helps establish whether neurofeedback is superior to, or at least non-inferior to, standard treatments.

Example:

‘Potential participants are screened for eligibility, and those who are eligible are randomly assigned to the treatment group (receiving rtfMRI NFT in addition to treatment as usual) or the control group (receiving only treatment as usual).’ (Cox et al., 2016).

**Control measures**

**Item 3a. Collect data on psychosocial factors**

This item is encouraged.

For example, participant motivation, treatment expectation, effort exerted, and subjective sense of success.

Examples:

(i) ‘To compare the NFT and the pseudo NFT group concerning the plausibility of the intervention, a subject self-report was utilized. Subjects reported on motivation to participate in the study, commitment to the study (before each session), and difficulty of the session (right after each session) using a seven-point Likert-scale (1 = not at all to 7 = very strong).’ (Enriquez-Geppert et al., 2014).

(ii) ‘In the present study, the effects of sex of participant, sex of experimenter, as well as the role of locus of control in dealing with technology will be investigated … Although the purpose of the present study is not to investigate further the effects of mindfulness and SMR baseline power on neurofeedback training outcomes, their impact will be measured and controlled statistically in the experimental design.’ (Wood and Kober, 2018).

**Item 3b. Report whether participants were provided with a strategy**

This item is essential.

If strategies were provided, report the details of the strategies.

Examples:

(i) ‘Importantly, the experimenter did not provide any explicit instruction to the participant regarding strategies; rather participants were told to increase the number of counts and bell rings by any mental means they could.’ (Davelaar et al., 2018).

(ii) ‘Subjects were instructed to execute or imagine the kinesthetic experience of a sequential finger tapping task (index-middle-ring-little-index-middle-ring-little) from the first
Comment: Currently there is no standard regarding the provision of strategies, nor is there systematic research on which strategies are the most effective (see section ‘provision of strategies’ from Enriquez-Geppert et al., 2017). Motor-imagery-assisted brain-computer interface (BCI) is the exception.

Item 3c. Report the strategies participants used
This item is encouraged.
Examples:
(i) ‘The reported mental strategies and the subsequent categorization process are described in Table A1 of the Appendix in more detail.’ (Kober et al., 2013)
(ii) ‘Among them, the most efficient strategies were friends (1.625), love (1.4) and family (1.1) while the worst were anger (−2.0) and calculation (−0.15). The effects of some positive strategy subtypes like love (lover (1.67)), nature (hometown (1.5)) and family (brothers (2.0)) stood out.’ (Nan et al., 2012).

Item 3d. Report methods used for online data processing and artefact correction
This item is essential.
For example, detection and rejection/correction of ocular and muscular artefacts (EEG, MEG), and of cardio-respiratory and movement artefacts (functional MRI).
Examples:
(i) ‘Before the start-baseline measurement, an EOG calibration method (3 min) was implemented that calculates the subject-specific, artifact-associated frequency band. This was used for all following measurements for eye blink detection and rejection during further measurements (for details see Huster et al., 2014) … Thus, the subject-specific artifact-associated frequency band that was calculated in the EOG calibration measure was monitored. Whenever the mean amplitudes of a 2 s segment was higher than the subject-specific artifact-associated frequency band (minus one standard deviation), the segment was rejected and not used for feedback.’ (Enriquez-Geppert et al., 2014).
(ii) ‘Pre-processing of single-subject fMRI data included correction of cardiorespiratory artifacts using AFNI implementation of the RETROICOR method. The cardiac and respiratory waveforms recorded simultaneously during each fMRI run were used to generate the cardiac and respiratory phase time series for the RETROICOR.’ (Young et al., 2014).

Item 3e. Report condition and group effects for artefacts
This item is encouraged.
Report condition and group effects for the artefacts detailed for Item 3d (to test whether artefacts are more prevalent in certain participants and conditions).
Examples:
(i) ‘We observed an intra-subject effect of regulation condition on HR [heart rate] (F(2,52) = 6.092; p = 0.004), which was driven by an increased HR during the active (“UP” and “DOWN”) regulation conditions (Figure 6A). The relative difference between “UP” and “DOWN” conditions was not correlated with regulation capacity (2-tailed Pearson R = 0.038, p = 0.853, Figure 6C). For RVT [respiration volume per time], there was a trend for an intra-subject effect of regulation condition (F(2,52) = 3.148; p = 0.051, Figure 6B). Additionally, we found a correlation between the relative RVT-difference between the “UP” and “DOWN” conditions and regulation capacity (2-tailed Pearson R = −0.450, p = 0.018, Figure 6D).’ (Marxen et al., 2016).
(ii) ‘In Fig. 6, mean heart and breathing rates obtained during the different feedback conditions are plotted jointly for P02–P05 and P09 (with all values being in the normal range). While observed differences in heart rate across target-level conditions were extremely weak, slightly augmented breathing frequencies were detected for higher target-level conditions on a descriptive level.’ (Sorger et al., 2018).

Feedback specifications

Item 4a. Report how the online feature extraction was defined
This item is essential.
For example, a frequency band, frequency band ratio, single region of interest, or functional connectivity measure. Was it individualized or fixed across all participants? How was it extracted (e.g. number and location of electrodes)?
Examples:
(i) ‘In each session, the IAF [individual alpha frequency] was calculated as the peak frequency of the alpha band during the first base rate and UA [upper alpha] was defined as the frequency band from IAF to IAF+2 Hz.’ (Zoefel et al., 2011).
(ii) ‘For the localizer scan, real-time statistical analyses were carried out via an incremental general linear model (GLM) using Turbo-BrainVoyager (TBV) … Target ROIs in the respective groups were identified during a localizer scan based on the t-statistic of the contrasts of interest, which were defined as positive vs. neutral pictures in the NFE group and scene vs. face pictures in the NFS group. Target ROIs in the NFE group were limited to limbic and fronto-parietal portions of the anterior cerebrum based on models of emotion processing in the human brain [19].’ (Mehler et al., 2018).

Item 4b. Report and justify the reinforcement schedule
This item is essential.
For example, justify the reinforcement schedule, or the feedback threshold criteria, in relation to existing neurofeedback literature and practice. Report how the feedback was given (e.g. continuous or periodic, proportional or binary). Report the amount of reward (e.g. percentage) per subject and across subjects.
Example:
Item 4c. Report the feedback modality and content
This item is essential.
Identify the feedback modality (e.g. visual, auditory, tactile, proprioceptive), and the feedback format (e.g. video clip, simple graphic, melody, tone).
Example:
‘Children from one group received the NFB treatment using as reinforcement an auditory stimulus (Auditory Group, AG), and children of the other group received a NFB treatment using as reinforcement a visual stimulus (Visual Group, VG) … The auditory stimulus was a tone of 500 Hz at 60 dB, and the visual stimulus was a white square of 20 cm² over a black background of a computer monitor.’ (Fernández et al., 2016)

Item 4d. Collect and report all brain activity variable(s) and/or contrasts used for feedback, as displayed to experimental participants
This item is essential for points (ii) and (iii); and we encourage researchers to include points (i) and (iv–vi).
Time points may include: (i) a pre-training baseline; (ii) rest blocks; (iii) training blocks; (iv) a post-training baseline; (v) transfer run(s) without neurofeedback; and (vi) long-term follow-up. Report the relevant units.
Example:
‘Thus the aim of this study was to focus on alpha neurofeedback and examine changes in three different measures: amplitude, percent time, and integrated alpha, across four methods: within sessions, across sessions, within sessions compared to baseline, and across sessions compared to baseline.’ (Dempster and Vernon, 2009).

Item 4e. Report the hardware and software used
This item is essential.
Include the versions.

Outcome measures (brain)

Item 5a. Report neurofeedback regulation success based on the feedback signal
This item is essential.
Identify the baseline or contrast used (e.g. subject-specific data from a previous session, reference data based on averaged data from a normative group). Identify the comparator run (e.g. training run or transfer run). Report both statistically significant and non-statistically significant findings.

Comment: We raise this point because some experiments report only the changes in a subset of brain activity that was not used for the neurofeedback signal.

Item 5b. Plot within-session and between-session regulation blocks of feedback variable(s), as well as pre-to-post resting baselines or contrasts
This item is essential.
Plotting the session course by comparing the session beginning, middle, and end (for instance, by arbitrarily dividing sessions into segments or using session blocks) allows the assessment of within-session dynamics. Between-session comparisons allow the assessment of the whole training course on a temporally more abstract level.
Example:
‘Thus, relative to the VC group, the VTA feedback group showed enhanced activation over the duration of the ACTIVATE trial … Relative to baseline, the VTA Feedback group increased activation in the first half of the trial (t(18) = 4.74, p < 0.0005) … In addition to group differences, VTA Feedback group activation at Post-test was significantly greater than Pre-test (t(18) = 2.36, p < 0.05) and greater than baseline (early: t(18) = 2.88, p < 0.05; late: t(18) = 3.29, p < 0.005; overall: t(18) = 3.52, p < 0.005).’ Also, see Fig. 3 in MacInnes et al. (2016).

Item 5c. Statistically compare the experimental condition/group to the control condition(s)/group(s) (not only each group to baseline measures)
This item is essential.
Comparing experimental and control groups/conditions to their respective baselines, but not to each other fails to test whether the experimental intervention outperforms the control intervention(s) (Nieuwenhuis et al., 2011).
Example:
‘Figure 2 … Amygdalar hemodynamic response was assessed using fMRI during exposure to (A) masked sad face presentations (SN-NN condition) and (B) masked happy face presentations (HN-NN condition). Error bars indicate ±1 SEM. * indicates a significant difference from the corresponding baseline at p corrected < .05. # indicates a significant difference from the experimental group at p corrected < .05.’ (Young et al., 2017a).

Outcome measures (behaviour)

Item 6a. Include measures of clinical or behavioural significance, defined a priori, and describe whether they were reached
This item is essential.
For example, by using MCIDs to establish the magnitude of an effect to interpret as clinically meaningful (see Engel et al., 2018; Lakens et al., 2018 for an overview on establishing MCID values and smallest effect sizes of interest). Many of these values remain open to discussion—explain the reasoning behind the value used. Moreover, collect data on acceptability, safety, and adverse effects.
In this paper, we are using the term ‘behaviour’ in the broad sense to encompass all non-physiological measures, including self-reports. Examples:

(i) ‘Minimal clinically important differences (MCIDs) were defined as “the smallest differences in scores in the domain of interest, which patients perceive as beneficial, and which would mandate, in the absence of troublesome side effects and excessive costs, a change in the patient’s management”... The MCID value for the 10-m walk test was 0.19 m/s; \(^{45}\) 3.5 s for TUG; 46 and 5 points each for the UPDRS-Brad and UPDRS-III.\(^{47}\) The MCID values of 5 points and 2 points were adopted for BBS and PDQ-39 (mobility), respectively.\(^{45,48}\) (Costa-Ribeiro et al., 2017).

<table>
<thead>
<tr>
<th>Domain</th>
<th>Item #</th>
<th>Checklist item</th>
<th>Reported on page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-experiment</td>
<td>1a</td>
<td>Pre-register experimental protocol and planned analyses</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>Justify sample size</td>
<td></td>
</tr>
<tr>
<td>Control groups</td>
<td>2a</td>
<td>Employ control group(s) or control condition(s)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>When leveraging experimental designs where a double-blind is possible, use a</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>double-blind</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td>Blind those who rate the outcomes, and when possible, the statisticians involved</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2d</td>
<td>Examine to what extent participants and experimenters remain blinded</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2e</td>
<td>In clinical efficacy studies, employ a standard-of-care intervention group as a</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>benchmark for improvement</td>
<td></td>
</tr>
<tr>
<td>Control measures</td>
<td>3a</td>
<td>Collect data on psychosocial factors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3b</td>
<td>Report whether participants were provided with a strategy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3c</td>
<td>Report the strategies participants used</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3d</td>
<td>Report methods used for online-data processing and artefact correction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3e</td>
<td>Report condition and group effects for artefacts</td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>4a</td>
<td>Report how the online-feature extraction was defined</td>
<td></td>
</tr>
<tr>
<td>specifications</td>
<td>4b</td>
<td>Report and justify the reinforcement schedule</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4c</td>
<td>Report the feedback modality and content</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4d</td>
<td>Collect and report all brain activity variable(s) and/or contrasts used for feedback, as displayed to experimental participants</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4e</td>
<td>Report the hardware and software used</td>
<td></td>
</tr>
<tr>
<td>Outcome measures</td>
<td>5a</td>
<td>Report neurofeedback regulation success based on the feedback signal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5b</td>
<td>Plot within-session and between-session regulation blocks of feedback variable(s), as well as pre-to-post resting baselines or contrasts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5c</td>
<td>Statistically compare the experimental condition/group to the control condition(s)/group(s) (not only each group to baseline measures)</td>
<td></td>
</tr>
<tr>
<td>Behaviour</td>
<td>6a</td>
<td>Include measures of clinical or behavioural significance, defined \textit{a priori}, and describe whether they were reached</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6b</td>
<td>Run correlational analyses between regulation success and behavioural outcomes</td>
<td></td>
</tr>
<tr>
<td>Data storage</td>
<td>7a</td>
<td>Upload all materials, analysis scripts, code, and raw data used for analyses, as well as final values, to an open access data repository, when feasible</td>
<td></td>
</tr>
</tbody>
</table>
Item 6b. Run correlational analyses between regulation success and behavioural outcomes

This item is essential.

Examples:

(i) ‘For the mean alpha amplitude at P4 (the NFB controlled parameter), we found no significant correlations with any neglect severity measures (i.e. omissions on the left, center, or right parts of the cancellation test, deviation on line bisection). However, as shown in Table 2, for the alpha variability and its left-right parietal asymmetry, we observed significant correlations with performance on the cancellation test.’ (Ros et al., 2017).

(ii) ‘The exploratory robust regression analysis suggested that changes in self-efficacy predicted residualized depression scores at the primary endpoint (R² = 0.18, adjusted R² = 0.15, β = –0.187 ± 0.073, Fig. 2c), such that increase in self-efficacy was associated with less depression severity (t30 = –2.551, p = 0.016).’ (Mehler et al., 2018).

Data storage

Item 7a. Upload all materials, analysis scripts, code, and raw data used for analyses, as well as final values, to an open access data repository, when feasible

This item is encouraged.

Description of consensus process

The authors T.R., S.E-G., and R.T.T. developed the idea for a checklist of this type. They worked together, in the form of an adversarial collaboration, to produce an initial outline of the present checklist. They then requested input from researchers involved in recent correspondences on neurofeedback, particularly those published in Brain and Lancet Psychiatry. These researchers included K.D.Y., J.S.S., S.R.S., R.S., M.S., F.S., M.A.S., J.A.M-F., D.M.A.M., J.L., D.E.J.L., R.J.H., J.G., T.F., and M.A. T.R., S.E-G., and R.T.T. then worked together to implement the comments from the researcher listed above and produce a first complete draft. This first complete draft was then sent to neurofeedback researchers involved in relevant discussions at recent conferences [e.g. Society for Applied Neuroscience (SAN) 2016; real-time Functional Imaging and Neurofeedback (rtFIN) 2017; Journée Nationale sur le Neurofeedback 2018], as well as the first-round contributors, to ask: (i) whether they agreed with the contents of the checklist; (ii) whether they would like to add, modify, or remove any material; and (iii) to invite researchers they believe may be interested in joining or commenting on the consensus. Together, T.R., S.E-G., and R.T.T. discussed each of the second-round comments and implemented those they believed appropriate for this checklist. Not all comments were addressed; in particular, specific comments relevant to only a subset of neurofeedback research, as well as a few points where contributors disagreed, were excluded from the present checklist. This second draft was then shared with all contributors before submitting for publication.

Funding

No funding was received towards this work. R.T.T. is supported by a postdoctoral fellowship from the Fonds de la recherche en santé du Québec. A.O. and M.A.L. are supported by the Center for Biologic Interfaces National Research University Higher School of Economics, Russian Federation Government grant, ag. No.14.641.31.0003.

Competing interests

U.S. has been paid for public speaking by Novartis, Medice, NeuroCare, the German Society for Biofeedback, the German Society for Psychotherapy and Psychiatry and the Akademie König und Müller. K.R. has received a grant from Takeda for another project. M.H. has a patent application for fNIRS neurofeedback, titled ‘Methods and systems for treating a subject using fNIRS neurofeedback’ (PCT/US2017/036532, filed June 8, 2017) as well as a contract with Elsevier to edit a book titled ‘FMRI Neurofeedback’. D.B. serves as an unpaid scientific advisor for an EU-funded neurofeedback trial unrelated to the present work. B.B. was paid for public speaking by the neuroCare Group (München, Germany). M.A. is unpaid chairman of the Brainclinics Foundation, a minority shareholder in neuroCare Group (Munich, Germany), and a co-inventor on four patent applications related to EEG, neuromodulation and psychophysiology, but receives no royalties related to these patents; Research Institute Brainclinics received research funding from Brain Resource (Sydney, Australia), Urgotech (France) and neuroCare Group (München, Germany), and equipment support from Deymed, neuroConn, Brainsway and Magventure. R.T.T. has received payments to consult with neurofeedback start-up companies. R.C.D. holds patents related to rtFMRI and rtFMRI-based feedback, and is CEO and a shareholder in Omneuron, a company that has developed technology related rtFMRI-based feedback. All other authors report no competing interests.

Supplementary material

Supplementary material is available at Brain online.

References

Algermissen J, Mehler DM. May the power be with you: are there highly powered studies in neuroscience, and how can we get more of them? J Neurophysiol 2018; 119: 2114–7.


Schabus M. Reply: on assessing neurofeedback effects: should double-blind replace neurophysiological mechanisms? Brain 2017; 140: e64.


Downloaded from https://academic.oup.com/brain/article-abstract/143/1674/1685/5807912 by guest on 03 August 2020


