The Efficacy of Neurofeedback and Traditional Therapies with Autism Spectrum Disorder in a Clinic Setting

Alexander Ryan[a*], Nicole Files[b]

[a]*PhD (Psych), Good Start Psychology, South Australia*
[b]*BPsych (Hons), Good Start Psychology, South Australia*
Email: [a]*alexryan@goodstartpsych.com.au*, [b]*nicolefiles@goodstartpsych.com.au*

Abstract

The efficacy of amplitude neurofeedback (NFB) in a clinic-based setting was explored in this study using a sample of 132 participants. 96% of participants were under 21 years of age. The sample appeared to be representative of the wider autistic child population. The current study used an *ex post facto*, quasi-experimental design. Using a clinically meaningful measure of change, called the Goal Attainment Scale, NFB participants demonstrated moderate gains on a targeted, challenging behavior or emotional problem. These gains were equivalent to a clinical response (> 50% reduction in symptoms) in traditional research terms. The NFB group performed better in absolute terms compared with traditional therapies and with NFB combined with traditional therapies. These group differences were not significant. A number of other factors did not influence clinical outcome (e.g., age, diagnosis, protocol type) suggesting that assessment-based NFB protocols, informed by experienced NFB practitioners, and delivered by NFB practitioners with two years' experience or less, can translate into significant and meaningful improvements for people on the spectrum. The limitations of this study were addressed.

Keywords: EEG; amplitude; neurofeedback; neurotherapy; autism spectrum disorder; ASD; comparison; traditional therapies; ex post facto study; clinic setting; children.

ARTICLE INFO		
RECEIVED	August 30, 2019	
REVIEWED	October 24, 2019	
ACCEPTED	November 6, 2019	

* *Corresponding author.*

Email: alexryan@goodstartpsych.com.au

1. Introduction

The efficacy of amplitude neurofeedback (NFB) with autism spectrum disorder (ASD), beyond case study analysis, has been reported in over twenty research trials since Jarusiewicz's (2002) breakthrough publication. All studies have demonstrated a positive effect, ostensibly a specific effect. Many of these studies are cited in a literature review by Perl and Perl (2019). There is also a home-based study on the auditory-only feedback system called Mente, which reported a positive effect on an autistic group, relative to a sham group (Carrick, Pagnacco, Hankir, Adbulrahman, Zaman, Kalambaheti, Barton, Link & Oggero, 2018). More recent studies are demonstrating that specific training protocols (C4-A2) are translating into brain activation associated with social functioning, reduced symptom severity and improvements in social behaviors on standardized measures (Datko, Pineda & Muller, 2018).

Despite Jarusiewicz's randomized controlled design (showing a positive effect of NFB compared to a control group), there is still no study that compares NFB with a traditional-therapy control group. Moreover, there is only one study with a sample size exceeding 30 subjects per experimental group (Thompson, Thompson & Reid, 2010). Thompson et al. reported a post-hoc analysis of 150 cases using a test-retest design and showed significant improvements in clinic-based participants across several variables including objective measures of attention (e.g. TOVA), core autistic symptoms, achievement (Wide Range Achievement Test) and intelligence (WAIS).

Despite these studies, a debate is brewing in the literature. On the one hand, literature reviews (e.g. Coben, Linden and Myers, 2010) critically analyze the research as suffering small sample sizes, non-random assignment, non-blinded participants and insufficient control of placebo (or contextual / nonspecific) effects. On the other hand, there are others who challenge the necessity for scientific rigour. Moshe and Moshe (2019) wrote that "clinic studies with controls eliminate some of the non-specific/unspecified elements (some say placebo) of the intervention as the cause of change. We need studies that are done with controls in a clinical setting for several reasons. Firstly, as clinicians, we do not seek to eliminate non-specific effects, we seek to maximise them. Furthermore, we do not randomly accept clients – they choose us, so the issue of random assignment to treatments is contrary to clinical practice. In addition, clients in clinic studies reflect real world conditions and comorbidities – the same as that which walks through our doors."

As a support of the latter argument, La Vaque et al. (2002), a task force who outline guidelines on the efficacy of psychophysiological interventions, write that several meta-analyses found no difference between retrospective, observational studies and randomized controlled experiments, suggesting that a study like ours may approach the validity of a gold standard methodology e.g., randomization, sham control. This comparability in efficacy is most likely in physician-delivered therapies (Benson & Hartz, 2000), rather than lifestyle-orientated research like diet, exercise or lifestyle change.

Based on the second school of thought, combined with a lack of evidence comparing NFB with traditional therapies, and a need for larger sample comparisons, we explored amplitude NFB with traditional therapies for ASD participants who frequented a clinic for therapies tackling the core problems of autism. We sought to compare: a) NFB with and without other therapies, and with traditional therapies alone, and b) ASD with or without comorbidities, and with other developmental disorders. The aim was to analyse the efficacy of NFB compared to evidence-based traditional therapies using a larger sample size than found in the current literature. This is an *ex post facto*, quasi experimental design.

2. Methodology

2.1. Participants

A total sample size of 132 clients were drawn from four practitioners at Good Start Psychology during the months of November and December, 2018. One practitioner (the author) has two years of experience with NFB and 20 years of clinical experience in the psychology field, whilst the three other practitioners were provisional psychologists with less than 12 months of experience each in a clinical role. Of the three Provisionals, two were female, one was male. The least amount of cases added from one practitioner was 27 and the most added was 36. There was a relatively even contribution of cases from each practitioner.

The only exclusion criterion was clients who attended less than 6 sessions (n=31, 19%). These clients were excluded because they had not engaged in therapy sufficiently to warrant a reliable analysis of the efficacy of therapy. These six sessions included one to two assessment sessions. In all, participants of this study had completed at least four sessions of a specific therapy or therapies.

All participants, or their primary caregiver, signed a consent form at intake to allow the publication of these data, in a de-identified group format, for the purposes of an article of this nature.

2.2. Data collection

Outcome data was measured using the Goal Attainment Scale (GAS) method, which was developed with chronic, mental conditions (e.g. persistent pain) as a person-centered measure of functional capacity and improvement. The GAS is developed on a five-point scale with a measure recorded at intake and then reviewed periodically to assess how the participant is progressing with a targeted, functional behavior. Criticism has been levelled at the GAS method for its untested assumptions of scaling (i.e. interval versus ordinal),

inter-individual equivalence and comparability, and reliability of coding across different behavioral observation measures (Ruble, McGrew & Toland, 2012). Ruble et al. (2012) demonstrated that GAS descriptions were evaluated for equivalency, behavior samples are representative, and different sources of behavior samples can be reliably coded. They recommended practicing writing GAS templates, and using standardized approaches in writing these goals.

The first author delivered training and weekly supervision to Good Start practitioners on how to write goals using the GAS method using templated examples. Frequent review and feedback was provided to training practitioners with writing these goals to ensure uniformity and consistency. Guidelines were also developed including: a) observable behaviors, where possible; b) two to three factors represented only; and c) five levels of change i.e., -1 = regression, 0 = intake functioning, 1 = mild improvement, 2 = moderate improvement (approximately a 50% reduction in symptoms) and 3 = large improvement (usually infrequent or complete absence of challenging state or behavior).

With the assistance of the primary care giver, the key presenting problem or challenging behavior was identified at intake and then rated on a five-point scale from -1 through to +3. A typical example of a GAS is as follows:

Overall goal: reduction in meltdowns.
-1 (regression): Greater than 4 meltdowns a day, intensity of 6 out of 10, lasting greater than 30 minutes per episode.
0 (intake/baseline): 4 meltdowns a day, intensity of 5 out of 10, lasting 30 minutes per episode.
+1 (slight improvement): 1 meltdown a day, intensity of 4 out of 10, lasting 20 minutes per episode.
+2 (moderate improvement): 1 meltdown a week, intensity of 3 out of 10, lasting 10 minutes per episode.
+3 (large improvement): 1 meltdown less than once per week (occasional frequency), intensity of 3 out of 10, lasting less than 10 minutes per episode.

The GAS example shows that there are three factors represented i.e., number of meltdowns, intensity and episode length. It is important to note that this means that, unlike a standardized measure or scale, the same score on the GAS could mean something quantitatively different from one client to the next. The GAS is therefore meaningful to each participant and sensitive to change on a targeted behavior or emotional problem.

When completing our outcome data analysis, the baseline GAS was zero for all cases. When the target behavior was reassessed, it was elicited by the practitioner from the participant and caregiver, and the GAS rating was evaluated by the principal psychologist (first author) with the practitioner to calibrate the rating to ensure it was consistent with self-reports. We did this for all cases other than the first authors' cases. There were a few cases when these data (n=12) were changed; in all cases, the rating either increased or decreased by 1-point only with no trend towards inflating or deflating the outcome.

In order to ensure that the GAS was relatively consistent and standardized, the first author's GAS scores were correlated with the second author's. To put it into context, Nicole Files does not work with participants, and is a research psychologist, and she was not provided any training or supervision in the GAS method. She merely reviewed each case's notes, examined the GAS and then estimated a reasonable follow-up rating based on caregiver's reports of participant progress with the target behavior.

Cohen's kappa (Cohen, 1960) coefficient was calculated to determine the inter-rater reliability between the first and second authors' GAS scores. The interrater reliability for the two raters' GAS scores was found to be kappa = 0.69 ($p<.001$), which according to guidelines from Landis and Koch (1977) can be described as substantial in strength.

2.3. Neurofeedback training

All participants were trained on the EEGer software application using A202 or J202 amplifiers. Most of our participants were children and the most common method of feedback used in this subgroup was DVD game. In this method of feedback, the participant watches a movie (of his/her choosing) and two types of feedback are provided including a fading of the screen and a beeping noise. A clear screen and beeping are contemporaneously presented (usually .8 to 1 second delay) when the participant's EEG data satisfies practitioner-determined parameters on a single reward and two or three inhibit band protocol, depending on the predetermined protocol. The total rewards for the session were generally maintained between 30-60% of the time to entrain the brain to function better by rewarding around a third to one half of the time. The logic is that we are giving the brain feedback that it is doing well but could function better. A criticism of some past studies is that they have rewarded 60 to 80% of the time which suggests the brain is rewarded highly for poor activity. Sessions ran for approximately 30 minutes and most participants were trained once per week only.

2.4. Choosing protocols

Training protocols were determined by recording EEG "minimaps" for each participant for one-minute at Fz, Cz and Pz, during eyes open and eyes closed phases. These data were reviewed by the first author, who also received regular mentoring from Dr Moshe Perl (20+ years of experience with NFB) around montages and protocols using EEGer with ASD cases. EEG-assessment guided NFB performs better than symptom-derived NFB training (Coben & Myers, 2009).

Dr Perl reviewed most of the minimaps recorded by the practitioners in our clinic. A consistent finding with ASD cases using minimaps is an over aroused brain (excess slow wave, usually delta and theta at all three sites), sometimes an over activated brain (excess fast wave or high beta) and lower than normal levels of beta wave activity.

Protocols are derived by a clinical process of examining the presenting symptoms and behavioral problems, and then examining the EEG minimap data. To elaborate on clinical process, if a participant was experiencing meltdowns, with elevations in delta and theta at Fz, we routinely trained T4-F4 (or C4-F4 for a child under 10 years of age); if the participant was complaining of sensory overwhelm or sensitivities, with elevations in the parietal cortex, we trained T4-P4; and consistent with evidence-based practice, if an EEG showed fronto-central slowing, with concomitant attentional deficits, we trained Fz-A1 or Cz-A1. In most cases, training involved calming the right side of the brain, derived from an arousal-regulation model of EEG biofeedback.

3. Results

Table 1 shows that there was a three to one ratio of males to females in our sample, which is consistent with the wider autistic population averages. The mean age was 13.3 (median age of 12 years) with an age range of 5 to 64 years. 127 (96%) participants were 20 years or younger.

The mean number of therapy sessions was 23. The range of sessions was 6 to 70. Table 1 shows that nearly half the sample (48.5%) engaged with NFB only. Non-NFB therapies included CBT (n=27, 38%), Play Therapy (n=10, 15%), Psych-education (n=15, 22%), Social Skills (n=10, 15%) and ABA therapy (n=6, 9%).

A significant proportion of the sample were formally diagnosed with ASD and a comorbidity (25.8%). These additional diagnoses included ADHD, ID, OCD, gender dysphoria and anxiety. The remaining non-ASD cases (n=25) were diagnosed with developmental conditions like ID, global developmental delay and Down syndrome, and acted as an additional comparison group.

A large proportion of the sample reported and targeted a challenging behavior of meltdowns, anger, aggression or emotional regulation problems (59%). A smaller proportion of cases identified anxiety (n=29, 22%) as the main presenting problem, and then social skills (n=8, 6%). The remaining cases focused on attentional difficulties, depression, toileting and migraines (n=9, 13%).

Finally, we divided NFB protocols into three main groups for our data analysis, namely, 1) T4-F4, 2) training on the sensory motor strip (SMS) and 3) a collection of other protocols (e.g. Fz-A1, T4-P4).

Table 1. Descriptive data of total sample

Variable Group	Sample size	Percentage
Gender		
Male	99	75%
Female	33	25%
Type of therapy		
NFB only	64	48.5%
NFB and therapy	35	26.5%
Therapy only	33	25%
Diagnosis		
ASD	73	55.3%
ASD plus comorbidities	34	25.8%
Non-ASD	25	18.9%
NFB Protocol		
T4-F4, C4 -F4	42	31.8%
SMS (Sensory motor strip)	28	21.2%
Other	27	20.5%
Main problem		
Melt downs / emotional regulation	56	42.4%
Anger / Aggression	22	16.7%
Anxiety	29	22%
Social skills	8	6.1%
Other	17	12.9%

The mean improvement across the 132 cases on the GAS was 1.8 and the median improvement was 2. The lowest score was 0 and the highest score was 3. No cases were rated as regressing through therapy i.e. rated -1. Fourteen cases rated no change, 34 rated a slight improvement, 48 rated a moderate improvement and 36 rated a major improvement. We did not complete a pre- and post-test analysis of scores because the pre-score on the GAS is always 0. That is, there is no variation at baseline.

Males (M=1.79, SD=.95) and females (M=1.85, SD=.96) did not show significant differences in outcome (F=.098, p=.76).

We then examined outcome ratings with factors such as therapy type, diagnosis group, presenting problem and NFB protocol. Figure 1 shows that the mean scores for each therapy group were: NFB = 1.92 (SD= .95), NFB+Therapy = 1.80 (SD= .93) and Therapy = 1.58 (SD= 1.00). Group differences were not significant (F=1.43, p=.245).

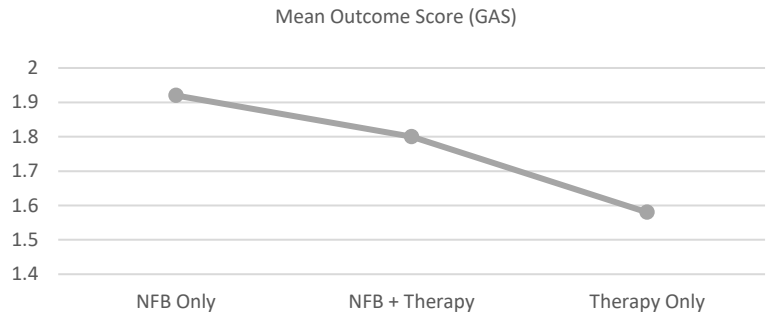

Fig.1. Outcome scores (GAS) by therapy type

Figure 2 shows that the diagnostic groups were even less discrepant in mean scores: ASD = 1.82 (SD= 1.00), ASD+Co-morbidity =1.71 (SD= .91), and Non-ASD = 1.88 (SD=.93). The differences were not significant (F=.265, p=.767).

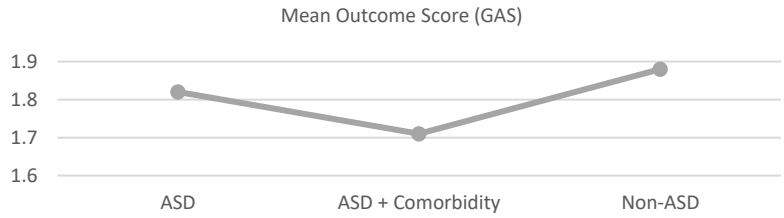

Fig.2. Outcome scores (GAS) by diagnosis group

Presenting problem and GAS means are represented in Figure 3. The mean scores were: Melt downs/emotional regulation = 1.80 (SD=.90), Anger/aggression = 1.91 (SD=1.02), Anxiety =1.72 (SD=1.00), Social skills = 1.63 (SD= .92) and other = 1.88 (SD=1.11). The differences between groups were not significant (F=.21, p= .933).

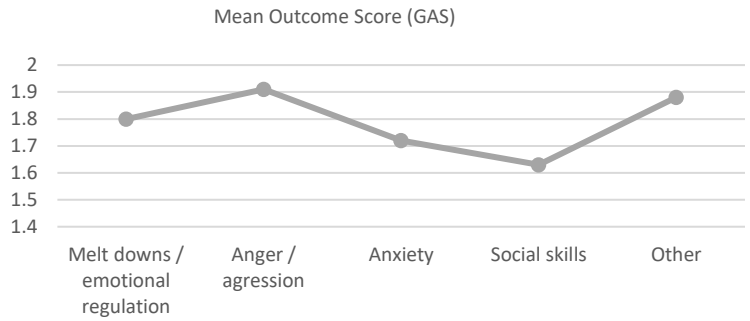

Fig.3. Outcome scores (GAS) by presenting problem

Finally, Figure 4 shows that NFB protocol groups produced the following GAS scores: T4-F4 = 1.71 (SD= .94), SMS = 2.04 (SD= .92) and Other = 2.04 (SD= .85). The differences were not significant (F= 2.10, p= .224).

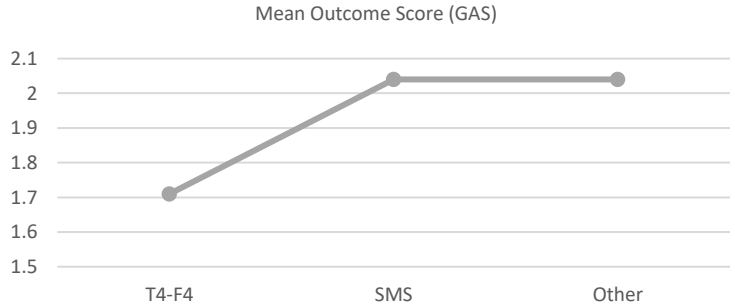

Fig.4. Outcome scores (GAS) by NFB protocol

There were no interaction effects using two-way ANOVAs for all independent variables based on outcome. Total number of sessions showed a significant positive relationship with outcome scores, $r(132) = .33$, $p < .01$. Age showed a negative relationship with outcome scores ($r(132) = -.05$, n.s.), but was not significant.

4. Discussion and Concluding remarks

The efficacy of NFB has been demonstrated in over 20 published articles with a home-based approach (i.e. Mente) also showing a positive effect. This study adds to the evidence-base by comparing NFB with and without other traditional therapies, with a larger sample size, which has not been demonstrated in the extant literature until now.

In absolute terms, the NFB-only group showed larger improvements than NFB with therapy and therapy alone. However, these group differences were not significant and we conclude that NFB performs as well as traditional, evidence-based therapies. Combining NFB and traditional therapies did not increase the outcome effect.

Given our sample of a three-to-one ratio of males to females, these findings generalize to the wider autistic population, with no expected significant therapy gains for individuals with autism versus those with comorbidities, or diagnoses other than ASD. Essentially, NFB could work equally with males or females, younger or older children, with or without ASD. Increasing the number of sessions appeared to increase the outcome score, although the correlation was weak.

Moreover, our study supports the argument and evidence for assessment-guided NFB protocols; however unlike past research (which were based on qEEG data), we have used only minimap data, which involves small time samples at three points, Fz, Cz and Pz. The minimap data, in combination with the presenting problem or challenging behavior, informed the protocols we adopted.

We know from existing research (e.g. Datko et al., 2018) that there are some protocols that work for specific problems. For example, social activation can be achieved by training C4-A2. However, an extensive review of the current research in this field shows that there are some fundamental flaws in reporting including the reference used in the montage used in the NFB training i.e. Cz-? (not disclosed). In our opinion, research to date has been helpful in showing the efficacy of NFB but there has been poor progress in identifying protocols that work for specific problems or pathology.

Essentially, we have not derived our protocols from research; we have derived our protocols from "minimaps" (limited EEG data) and clinical knowledge, using protocols that have worked in a clinic setting, with commercial participants.

4.1. Limitations

The current study will not satisfy the research scientist with a desire to see findings based on methodological rigour. Yet, we are cognizant of meta-analyses that show that comparison research like ours generate data equivalent to randomized, controlled methods, even though there is a common assumption that these inflate effect sizes or lack validity (Benson & Hartz, 2000; Concato, Shah, Horwitz, 2000).

Nonetheless, these findings cannot be attributed solely to the interventions implemented because of the possibility of contextual and unexplained factors. The contributing methodological factors leading to this conclusion include non-randomization of participants, varied NFB training protocols in the groups receiving NFB, non-blinded participants and therapists, and the confounding contribution of other therapies implemented at the time of our therapies e.g. speech therapy, occupational therapy. We cannot explain away contextual factors except that they were probably equal across the three groups, and blinding participants might be easily done, however blinding a trained NFB practitioner is very difficult and involves significant techniques in "sham" methodology. The presentation of EEG to the practitioner during a sham training session is pre-recorded, unbeknownst to him or her, and the participant receives random reward. Anecdotally, presenting pre-recorded or real EEG to experienced practitioners often causes error in guessing the authenticity of the EEG, however it is not scientifically proven if "sham" NFB training fully blinds the practitioner that pre-recorded EEG is a live feed of the participant's EEG.

This study was based in a "real" clinical practice and represents traditional clinical routine. The participant attends an appointment, discusses a challenging problem or behavior (s)he wants to reduce or eliminate and (s)he collaborates with the therapist for a therapy type. The participant plays a role in the therapy (s)he would like to participate in, and, in that regard, these effects were relatively equal across the groups. All participants chose the therapy that (s)he would like to receive and most received a benefit from therapy to a moderate extent. These moderate improvements are equivalent to a 50% reduction in symptoms, therefore a response rate in traditional clinical research terms.

In our opinion, due to contextual factors (like expectancy effects) we believe that the effects in this study may be larger than produced in a randomized, controlled trial. Having said this, Begemann, Florisse, van Lutterveld, Kooyman and Sommer (2016) reported a meta-analysis of NFB and autism trials and concluded large effect sizes for NFB compared with wait list controls on general symptomatology. That is, our findings are probably not inflated and not skewed to a positive effect. It suggests that NFB in a real clinical context is (at least) moderately efficacious as traditional evidence-based therapies.

In fact, there are many positives to NFB in clinics where individuals with autism are attempting to improve quality of life and functioning. It is non-invasive with no side effects; there are no pre-requisite skills or education (like with CBT); the participant can be

verbal or nonverbal (the positive effect does not discriminate based on receptive and expressive abilities, in our opinion); and the improvements generalize to other areas of functioning not targeted in therapy.

In our opinion, these findings are not scientifically rigorous, and cannot rule out nonspecific effects or contaminating factors, however they probably generalize to clinics that focus on ASD cases of children and young adults, with assessment-guided NFB protocols, using practitioners with NFB experience ranging from less than 12 months to two years, guided by highly experienced EEG therapists. Future research could focus on: a) assessing this therapeutic model, versus other service delivery models of NFB, b) investigate the efficacy of NFB protocols with different presenting problems and behaviors of concern, and c) how strongly the number of NFB sessions impact on outcome.

References

Begemann, M., J.R. Florisse, E., Lutterveld, R., Kooyman, M., Sommer, I., 2016. Efficacy of EEG neurofeedback in psychiatry: A comprehensive overview and meta analysis. Translational Brain Rhythmicity. 1(1): 19-29.

Benson, K., Hartz, A. J. 2000. A comparison of observational studies and randomized, controlled trials. New England Journal of Medicine. 342(25). 1878-1886.

Carrick, F. R., Pagnacco, G., Hankir, A., Abdulrahman, M., Zaman, R., Kalambaheti, E. R., Oggero, E., 2018. The Treatment of Autism Spectrum Disorder With Auditory Neurofeedback: A Randomized Placebo Controlled Trial Using the Mente Autism Device. Front Neurol. 9: 537.

Coben, R., Linden, M., Myers, T. E. 2010. Neurofeedback for autistic spectrum disorder: a review of the literature. Appl Psychophysiol Biofeedback. 35(1): 83-105.

Coben, R., Myers, T. E. 2010. The relative efficacy of connectivity guided and symptom based EEG biofeedback for autistic disorders. Appl Psychophysiol Biofeedback. 35(1): 13-23.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and psychological measurement. 20(1): 37-46.

Concato, J., Shah, N., Horwitz, R. I. 2000. Randomized, controlled trials, observational studies, and the hierarch of research designs. New England Journal of Medicine. 342(25): 1887-92.

Datko, M., Pineda, J. A., Müller, R. A. 2018. Positive effects of neurofeedback on autism symptoms correlate with brain activation during imitation and observation. Eur J Neurosci. 47(6): 579-591.

Jarusiewicz, B. 2002. Efficacy of Neurofeedback for Children in the Autistic Spectrum: A Pilot Study. Journal of Neurotherapy. 6(4): 39-49.

Landis, J. R., & Koch, G. G. 1977. The measurement of observer agreement for categorical data. Biometrics. 33(1): 159-174.

La Vaque, T. J., Hammond, D. C., Trudeau, D., Monastra, V., Perry, J., Lehrer, P. 2002. Template for developing guidelines for the evaluation of the clinical efficacy of psychophysiological interventions. Applied Psychophysiology and Biofeedback. 27(4): 273-281.

Perl, M., Perl, D. 2018. EEG amplitude neurofeedback: a review of the research. Asia Pacific Journal of Neurotherapy. 1(1): 44-55.

Ruble, L., McGrew, J. H., Toland, M. D. 2012. Goal attainment scaling as an outcome measure in randomized controlled trials of psychosocial interventions in autism. J Autism Dev Disord. 42(9): 1974-1983.

Thompson, L., Thompson, M., Reid, A. 2010. Neurofeedback outcomes in clients with Asperger's syndrome. Appl Psychophysiol Biofeedback. 35(1): 63-81.