Australian Journal of Educational & Developmental Psychology. Vol 9, 2008, pp 47-52

Critique

Amon, K., & Campbell, A.(2008). Can children with AD/HD learn relaxation and breathing techniques through biofeedback video games? *Australian Journal of Educational & Developmental Psychology,* Vol. 8, pp. 72-84.

Craig Wright[1]
Understanding Minds
Gold Coast, Australia
&
Elizabeth Conlon
Griffith University

In the 2008 volume of this Journal Amon and Campbell reported a successful trial of a commercially available biofeedback program, "The Wild Divine", in reducing symptoms of Attention-Deficit/Hyperactivity Disorder (ADHD) in a group of children with ADHD and a control group. They introduced their study by suggesting that side effects of medications, the efficacy (perhaps they mean "ethics") of medicating young children, and the possibility of future drug use means that research into non-medical interventions for ADHD is important in providing a greater array of treatment options. We applaud this notion; although we take exception to their latter justification. What little evidence there is on this subject suggests that using a stimulant reduces the chances of substance abuse in ADHD (see Barkley, Fischer, Smallish & Fletcher, 2003 for review and the meta-analysis of Wilens, Faraone, Biederman & Gunawardene, 2003). Amon and Campbell seem to have selected biofeedback as an experimental treatment on the basis of literature which suggests that cognitive or physiological techniques such as relaxation and meditation can lead to positive behavioural, emotional, and somatic outcomes for normal samples of children. They go on to suggest that the "relaxation" afforded by controlled breathing techniques contained in the experimental biofeedback game would lead to greater control over nervous system activity and hence performance in children with ADHD. No physiological measures are reported. In discussing their data, Amon and Campbell (2008) claimed: "the findings from this study show that the Wild Divine video game, as a biofeedback system, has the potential to produce positive developments on ADHD symptoms and disruptive behaviours, with few side effects" (pp.82). We contend that several methodological, reporting, statistical, and interpretative problems make this conclusion difficult to sustain. We will take each of these problems in turn.

Methodological and reporting criticisms

With regard to methodology, Amon and Campbell (2008) selected a treatment group of children with ADHD. This group was further divided into two sub-groups: one (*n* = 17) who received the treatment once a week and another (*n* = 7) who received the treatment more than

[1] Address for correspondence:
Dr. Craig Wright
Understanding Minds
PO Box 501
Mermaid Beach QLD
Australia 4218
craig@understandingminds.com.au
+61 7 55261516
Australia 4218

once per week. The latter either attended twice or three times according to the Amon and Campbell's Method. However, there is no report of how many of the 7 students attended twice or three times. The control group consisted of 12 children who were not diagnosed with ADHD; ten of whom attended training once a week, with two children attending more than once a week. Of the latter, there is again no mention of whether either attended twice or three times. We would like to see this information reported as it pertains to the intensity of the intervention. In addition, while Amon and Campbell recognise the difficulties created by unequal sample sizes, with n = 2 in a single group, the study has questionable external validity as two children are unlikely to be representative of performance within the population. Finally, there is insufficient power generated in a group of this size to made any conclusions from tests of statistical significance be they parametric or non-parametric.

The number of parents from both the ADHD and control groups was reported to be fewer than the number of children in the study. For example, there were 24 children in the ADHD group and only 19 parents. In the control group there were 12 children and 8 parents. These descriptive data suggest that the authors used more than one child from a single family. If true, this information should have been reported, as it influences the independence of the scores obtained for each child. In addition, if more than one child from a family participated, the scores obtained from a single parent may have had a large influence on the overall results, questioning the ability of the study to make general statements about the population.

Of the treatment groups, 62.5% were already taking a stimulant medication or atomoxetine. No attempts were made to determine whether the effects of the treatment differed for children on medication when compared to children who were not (or at least none was reported). Medication was however, used to explain unpleasant side-effects reported in the treatment program (p. 80).

Each participant, control and treatment, attended a University clinic where they engaged in the breathing/biofeedback treatment for 45 minutes per session. The treatment itself is not well described; certainly not to the point of being replicable even if purchasing the commercial software, and perhaps even then, given that Amon and Campbell (2008) state that a researcher guided the child through the game, helped when they needed direction, or "aided in motivating with the breathing technique required for an activity" (p 75). We would like more detail on the treatment, particularly on the role of the experimenter which does not seem part of the commercial "Wild Divine" program and which would therefore be vital in attempts to replicate.

Amon and Campbell also failed to report the range of scores for their pre-treatment questionnaires. They reported means and standard deviations which only allow the reader to *assume* that the scores obtained for each group represented a profile of children with or without ADHD. However, if ranges were to be reported, it would have provided the reader with some indication of the upper and lower limits of severity of the difficulties experienced by children in each group. Without ranges the reader is left with the uneasy feeling that the groups may not be as different as one might think. This difficulty is exacerbated by use of an unstandardised measure of ADHD which created further difficulty in interpreting these scores.

Finally, Amon and Campbell chose as outcome measures the Strengths and Difficulties Questionnaire (SDQ; which should be attributed to Goodman, 1997), an unstandardised ADHD questionnaire based on DSM-IV-TR (APA, 2000), and an unstandardised questionnaire that asked subjective questions of parents regarding their perception of how difficult and frustrating their child found the 'Wild Divine' game. The SDQ was an interesting inclusion as a general measure of psychopathology. However, given the lack of questions specific to ADHD it is not a measure that has validity when evaluating treatment effects on the core symptoms of ADHD, particularly when it appears that an overall score and not individual dimension scores were used.

The ADHD questionnaire was apparently based upon the DSM-IV-TR (APA, 2000). One might assume the measure to be valid given that a number of other similar measures are used operationalise the DSM-IV-TR symptoms using a 4-5 point Likert scale. Nevertheless, given the cheap (and in some cases free) and easy availability to standardised questionnaires that also have normative data (e.g., the Disruptive Behaviour Rating Scale; Barkley & Murphy, 2006), one wonders why the authors chose to develop and report an unstandardised measure for which no normative data were available. In addition, the questionnaire was not reported in an Appendix of the paper, making conclusions based on this measure even more difficult. Furthermore, ADHD is a condition that may be defined by inconsistency and by differing behaviour across situations. In future studies we suggest that measures of impairment, such as the School- and Home-Situations Questionnaires (Barkley & Murphy, 2006) or the Children's Impairment Rating Scale (e.g., Fabiano & Pelham, 2002), be used to investigate treatment effects on ratings of social, family, and academic impairment across settings.

The information provided by the Wild Divine questionnaire may have been interesting in providing a fuller description of the treatment. However, given its subjective and unstandardised nature and because the items do not relate to symptoms of ADHD, it should be discounted as evidence for treatment effects. Amon and Campbell's reporting of this questionnaire also reveals further methodological difficulties. They reported that the "majority of parents (54%) in the experimental group reported that their children did practice (*sic*) the breathing techniques they learned through the game, away from the sessions" (p. 79). Based on this information, it would seem impossible to conclude with certainty that it was the treatment that resulted in the subjective changes in parents' judgments about behaviour when it may have resulted from *some* parents' satisfaction with their children practising the breathing strategies in the home setting.

Statistical criticisms

There are a number of statistical criticisms associated with this paper. First, based on a large number of tests of statistical significance, the authors appear solely to base conclusions on findings of statistically significant effects (NHST; i.e., $p < .05$), a persistent problem which has plagued the psychological and education literature. Second, the way these tests were conducted produced contradictory findings, which in view of the third criticism make the results almost impossible to follow. The lack of clear description of the results with reports is the third criticism. Different descriptions of the same data were provided of the same statistical tests or outcomes within various sections of the manuscript. While some mean scores were reported these appeared sporadically and often failed to correspond to the statistical tests conducted. Finally, there was no evidence presented that provided any indication of the practical or clinical meaningfulness of these data either for the groups of children used, or for individuals within the ADHD groups. These criticisms will be discussed for each the SDQ and ADHD measures.

Criticism of NHST is typically twofold and dates back at least to Rozeboom (1960), Meehl (1967), and Cohen (1990; 1994). The first problem is that NHST is based on the assumption that the null hypothesis (H_O) is exactly true in the population; however, the H_O is always false in the population given sufficient power (Campbell, 2005; Cohen, 1994). The corollary of H_O always being false in the population is that finding a statistical significant effect is a trivial matter of simply having a large enough sample size (Kirk, 1996). Many prominent researchers have advocated that, along with means, the 95% confidence intervals should be reported. These are useful when comparing findings across studies and when evaluating stability (e.g., Cohen, 1994; Kirk, 1996; Rozeboom, 1960; Wilkinson & APA Task Force on Statistical Inference, 1999).

The second problem with NHST is that a significant *p*-value does not describe the magnitude of the effect or the practical significance of a result (e.g., Cohen, 1992; Vacha-Haase & Thompson, 2004). Thus, many studies report statistically significant results (i.e., $p < .05$), even though the magnitude of the effect has little practical value (Ives, 2003). To avoid this problem, recommendations have been made for inclusion of effect size data to measure the practical meaningfulness of studies in which *p* values are used (Wilkinson & APA Task Force on Statistical Inference, 1999).

By using the means, standard deviations, and group sizes presented sporadically by Amon and Campbell, only a small number of effect sizes could be estimated. Those that are relevant and calculable are included in the discussion of the individual measures (i.e., the SDQ and ADHD questionnaire). For future reference, at a bare minimum, the means and standard deviations for each effect should be reported (Wilkinson et al., 1999) so that readers can estimate the size of effects.

SDQ data

This section of the Results is confusing and difficult to follow. While many statistical tests are presented, description of these findings is minimal. First, a statistically significant difference between the ADHD and control groups was reported, which is expected given the different diagnoses of individuals in the groups. Based on Figure 1, which presented the mean of each group (no measure of error was reported) at pre- and post intervention, a statistically significant multivariate interaction was then reported between time and group. No explanation of this interaction was provided.

Further complicating interpretation is that when the ADHD group was divided into those who attended once weekly versus those who attended more frequently (effects not evaluated in the omnibus ANOVA), children who attended one session per week showed a significant reduction in SDQ scores from pre- to post- intervention, while children attending multiple sessions did not. Given that the intensity of any intervention might be considered to be an important moderator of effect size (i.e., the higher the intensity, the stronger the effect), this seems an odd result and is not explained by the authors. The authors confirmed these findings with non-parametric statistics. However, neither means, medians, nor standard deviations were presented. In the following paragraph, means and standard deviations were presented for attendance frequency for pre-intervention, (M = 26.82, SD = 5.18) and post intervention, (M = 25.00, SD = 8.14). However, to which groups these data refer is not reported; nor do these means correspond to those presented in Figure 1 (p.77).

Further multivariate results were reported that show there were no statistical differences when comparing attendance frequency for the children with ADHD. The conclusion reported at this point was that attendance at either single or multiple weekly sessions failed to influence outcome on the SDQ. These effects are not consistent with the different effects of single and multiple session attendance reported in the previous paragraph. Further in additional analysis, it was claimed by the authors that both the single- and multiple-times-weekly strategies produced statistically significant improvements on the SDQ, with the multiple sessions resulting in greater gains (p. 81).

Furthermore, the authors reported that neither group moved from their SDQ category (i.e., the group with ADHD were remained in the "abnormal" range and the control group remained classified as "borderline" based on the SDQ conventions). Calculations of the effect sizes for the SDQ variable using Cohen's *d* based upon the reported means and standard deviations was .55 for the experimental group, which is considered moderate. The effect size from pre- to post-intervention for the control group produced a Cohen's *d* of .05, a negligible effect. While these effect size calculations suggest that some reduction in symptomatology was reported for the ADHD group, the clinical significance of these findings is unknown, as children with ADHD remained in the abnormal range of this general measure of symptoms of psychopathology. Together, these findings are certainly inconsistent with the reported interpretation given by the authors that "*both the experimental and control groups had significant reductions in the SDQ questionnaire, resulting in improvements in behaviour in the final session*" (p. 81).

In the Introduction, one of the important justifications for conducting the study was the reported effectiveness of relaxation training in normal functioning children. On this basis it would have been expected that there would have been some success in treatment for the psychological distress reported in the control group. Yet, there was no statistical significant effect for the control group; who despite being 'controls' fitted the SDQ descriptor of "borderline" thus indicating that at least some were experiencing emotional or behavioural difficulties. The authors make no attempt to explain this apparent anomaly.

ADHD questionnaire

The main statistical effect Amon and Campbell (2008) reported with regard to their ADHD questionnaire was a time × group interaction. In contrast to analysis using the SDQ, scores on the ADHD questionnaire were presented for pre-test, diary 2, diary 3 and diary 4, which were presumably obtained at one, two and three month intervals (although this does not appear to be explicitly reported). There are four measures presented in Figure 2 (p.78), which appear to describe the group × time interaction effect. Inspection of Figure 2 suggests some reduction in the parental reports of ADHD symptoms within the group with ADHD, but not for the control group. Multiple t-tests for the ADHD group were reported and a statistically significant reduction in symptoms of ADHD claimed between diaries 1-2 and 2-3, with no statistical differences between diaries 3 and 4. When changes across time were evaluated for the ADHD group who either attended multiple or single sessions, no treatment effect was found for children attending multiple sessions. However children attending single weekly sessions were reported to improve between diary 1 and 2, and diaries 2 and 3, but not between Diary 3 and 4 (p. 79). No descriptive statistics were presented and when further statistical analysis was conducted, Amon and Campbell reported that the number of sessions attended variable did not influence outcome (p. 79). The reader's confusion is justified as this conclusion contradicts the interpretation of outcome of the multiple t-tests reported in the preceding paragraphs. Finally, no significant changes were reported for the control group (p. 79), yet in the Discussion it was reported that significant improvements in scores on the ADHD questionnaire were obtained by the control group (p. 81).

In summary, Amon and Campbell (2008) presented a study with numerous methodological, statistical, and interpretative flaws. In carrying out evaluations of interventions, researchers have an obligation to conduct high quality studies and to present data with clarity. At the every least, the method should be replicable and the psychometric properties of the outcome measures fully described. Two influential papers on the reporting of research in psychology and intervention studies are the Wilkinson paper on statistical reporting in Psychology (Wilkinson & the Task Force on Statistical Inference, 1999) and Jacobson and Truax (1991). The latter emphasises the critical need to evaluate the clinical significance of the results in any intervention study. Both emphasise the importance of clear and accurate description of findings over reporting of multiple tests of statistical significance, which are in many cases impossible to interpret meaningfully.

The clinical meaningfulness of the measures used within a study is pertinent to the clinical significance of the study itself. In this study (Amon & Campbell, 2008) the ADHD questionnaire was unstandardised and the reader therefore has no way to determine the range of 'normal'. In the absence of this information one might assume that the control group represents "normality". Even if this were the case, and it is impossible to tell from the descriptive statistics presented, examination of Figure 2 (p.78) indicates that the ADHD group still scored substantially higher on the measure than the control group and were far from having "normal" scores. In addition, measures that constitute the amount of change required from pre- to post-intervention that indicate reliable change due to the treatment program (i.e., a reliable change index) and some measure of the clinical significance of the results must also be presented. Either a quantitative approach to clinical significance by determining the proportion of children in the ADHD group who return normal function following an intervention (e.g., Jacobson, Roberts, Berns, & McGlinchey, 1999) or a measure of the social impact of change (Kazdin, 2003) should be reported. This paper falls short of each of these criteria.

Our aim in writing this critique is simple. Although on the face of it unlikely to cause harm, alternative treatments for all developmental disorders may give a false sense that the symptoms and impairment are being addressed, thus delaying effective intervention. Clinicians reading this research report will be unable critically to evaluate the research results so would accept the analysis of the authors that this intervention has the potential to reduce psychological distress in both the ADHD and the control groups. The data do not support this conclusion. In addition, there are direct costs associated with the treatments as well as potential indirect costs such as loss of wages and time for working parents. In short, it should be clear that, while interesting, Amon and Campbell's data provide unconvincing evidence for biofeedback as an effective treatment for ADHD.

REFERENCES

American Psychiatric Association (APA). (2000). *Diagnostic and Statistical Manual of Mental Disorders- Text Revision* (4th ed.). Washington DC: American Psychiatric Association.

Amon, K. L., & Campbell, A. (2008). Can children with AD/HD learn relaxation and breathing techniques through biofeedback video games? *Australian Journal of Educational & Developmental Psychology, 8,* 72-84.

Barkley, R. A., Fischer, M., Smallish, L., & Fletcher, K. (2003). Does the treatment of attention-deficit/hyperactivity disorder with stimulants contribute to drug use/abuse?: A 13-year prospective study. *Pediatrics, 111(1),* 97-109.

Barkley, R. A., & Murphy, K. .R. (2006). *Attention-Deficit Hyperactivity Disorder: A Clinical Workbook* (3rd Edition). The Guildford Press: New York, NY.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45,* 1304-1312.

Cohen, J. (1992). Things I have learned (so far). In A.E. Kadzin (Ed). *Methodological issues and strategies in clinical research* (pp. 315-333). Washington: American Psychological Association.

Cohen, J. (1994). The earth is round (p <.05). *American Psychologist, 49*, 997-1003.

Fabiano, G.A., & Pelham, W.E. (2002). Measuring impairment in children with Attention-Deficit Hyperactivity Disorder. *The ADHD Report, 10 (2),* 6-10.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.

Ives, B. (2003). Effect size use in studies of learning disabilities. Journal of Learning Disabilities, 15, 490-511.

Jacobson, N.S., Roberts, L.S., Berns, S.B. & McGlinchey, J.B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology, 67,* 300-307.

Jacobson, N.S., & Truax, P. (1991). Clinical Significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12-19.

Kazdin, A. E. (1999). The meanings and measurement of clinical significance. *Journal of Consulting and Clinical Psychology, 67*, 332–339.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*, 746-759.

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science, 34*, 103-15.

Rozeboom, W. W. (1960). The fallacy of the null-hypothesis significance-test. *Psychological Bulletin, 57*, 416-28.

Wilens, T. E., Faraone, S. V., Biederman., J., & Gunawardene, S. (2003). Does stimulant therapy of attention-deficit/hyperactivity disorder beget later substance abuse? A meta-analytic review of the literature. *Pediatrics, 11(1),* 179-185.

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology, 51*, 473-481.

Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals. *American Psychologist, 54,* 594–604.

Biographical Note:

Dr Craig Wright is Principal Psychologist at Understanding Minds, a clinic that specialises in developmental and learning disorders. He is the author of the Understanding Words reading intervention program. *Dr Elizabeth Conlon* is Senior Lecturer at the Griffith University School of Psychology. She is known for her work on visual discomfort and research on the neurological factors that may underlie dyslexia.