Applied Psychophysiology and Biofeedback, Vol. 29, No. 2, June 2004 (© 2004)

The Effectiveness of Neurofeedback and Stimulant Drugs in Treating AD/HD: Part I. Review of Methodological Issues

Thomas Rossiter[1]

The paper examines major criticisms of AD/HD (Attention Deficit/Hyperactivity Disorder) neurofeedback research using T. R. Rossiter and T. J. La Vaque (1995) as an exemplar and discusses relevant aspects of research methodology. J. Lohr, S. Meunier, L. Parker, and J. P. Kline (2001), D. A. Waschbusch and G. P. Hill (2001), and J. P. Kline, C. N. Brann, and B. R. Loney (2002) criticized Rossiter and La Vaque for (1) using an active treatment control; (2) nonrandom assignment of patients; (3) provision of collateral treatments; (4) using nonstandardized and invalid assessment instruments; (5) providing artifact contaminated EEG feedback; and (6) conducting multiple non-alpha protected t tests. The criticisms, except those related to statistical analysis, are invalid or are not supported as presented by the authors. They are based on the critics' unsubstantiated opinions; require redefining Rossiter and La Vaque as an efficacy rather than an effectiveness study; or reflect a lack of familiarity with the research literature. However, there are broader issues to be considered. Specifically, what research methodology is appropriate for studies evaluating the effectiveness of neurofeedback and who should make that determination? The uncritical acceptance and implementation of models developed for psychotherapy, pharmacology, or medical research is premature and ill-advised. Neurofeedback researchers should develop models that are appropriate to the technology, treatment paradigms, and goals of neurofeedback outcome studies. They need to explain the rationale for their research methodology and defend their choices.

KEY WORDS: neurofeedback; AD/HD; stimulant drugs; research design; active treatment control.

INTRODUCTION

During the past decade, the use of neurofeedback (EEG biofeedback) to treat Attention Deficit/Hyperactivity Disorder (AD/HD) and other disorders has increased dramatically. Concurrently, neurofeedback has received growing attention from the public and mounting scrutiny and criticism from various professional communities, particularly those involved in behavior therapy and/or the treatment of AD/HD. Barkley (1992, 1993) and CHADD (Children and Adults with Attention-Deficit/Hyperactivity Disorder) were among the earliest and most vociferous critics of neurofeedback for AD/HD. More recent critics include

[1]1775 Highview Street, De Pere, Wisconsin 54115; e-mail: t.rossiter@worldnet.att.net.

1090-0586/04/0600-0095/0 © 2004 Plenum Publishing Corporation

Lohr, Meunier, Parker, and Kline (2001), Waschbusch and Hill (2001), and Kline, Brann, and Loney (2002), who reviewed the literature supporting the use of neurofeedback for AD/HD and other disorders and concluded that it is methodologically flawed and unconvincing.

The purpose of this paper is to examine major criticisms of AD/HD neurofeedback research using Rossiter and La Vaque (1995) as an exemplar and to discuss relevant aspects of research methodology. Lohr et al. (2001), Waschbusch and Hill (2001), and Kline et al. (2002) criticized Rossiter and La Vaque for (1) using an active treatment control rather than a waiting list or other nontreatment control; (2) the nonrandom assignment of patients to treatment and control groups; (3) providing collateral treatments to patients in each group; (4) using assessment instruments that were not standardized, lacked ecological validity, could not discriminate between AD/HD and other psychiatric disorders, and were not sensitive to treatment effects; (5) providing EEG feedback contaminated by ocular artifact; and (6) conducting multiple non-alpha protected *t* tests.

The issues raised by the critics relating to research design and methodology need to be addressed because some of the same criticisms can be directed at Rossiter (2003), Fuchs, Birbaumer, Lutzenberger, Gruzelier, and Kaiser (2003), and Monastra, Monastra, and George (2002). Each of these studies used an active treatment control, allowed patients to choose their treatment(s) rather than randomly assigning them, presumably provided artifact contaminated EEG feedback, and used the TOVA as the primary measure of treatment outcome. If the critics are correct, all of these studies are unsound and cannot provide evidence supporting the effectiveness of neurofeedback as a treatment for AD/HD. However, before accepting the criticisms of Lohr et al. (2001), Waschbusch and Hill (2001), and Kline et al. (2002) at face value, it would be prudent to first determine if they are valid and are supported by the research literature.

Rossiter and La Vaque (1995) used a quasi-experimental design and an active treatment comparison group (La Vaque & Rossiter, 2001) to evaluate the relative effectiveness of EEG biofeedback and stimulant drug programs in controlling the symptoms of AD/HD. They found that 20 sessions of a treatment program with EEG biofeedback as the primary component significantly reduced the symptoms of AD/HD in 19 of 23 patients. The stimulant drug group (methylphenidate, dextroamphetamine, or pemoline) matched on age, intelligence, diagnosis, and collateral treatments achieved similar results with 20 of 23 patients showing significant improvement. The treatment groups showed similar improvement on the Test of Variables of Attention (TOVA), a computer administered and scored continuous performance test (CPT) that provides an objective measure of the cognitive symptoms of AD/HD. Both groups improved significantly over pretreatment baselines on measures of attention (errors of omission), impulse control (errors of commission), processing speed (response time), and variability in attention (variability in response time). The percentage of patients improved in both groups (EEG biofeedback = 83%, stimulants = 87%) exceeded the highest placebo response rate (39%) reported in reviews of double blind placebo studies of stimulant drugs (Barkley, 1977). In addition to improvement in cognitive functioning on the TOVA, the neurofeedback group demonstrated parallel improvement in behavioral and emotional adjustment on the Behavior Assessment System for Children (BASC), a parent report questionnaire. Pre- and posttreatment results from the BASC indicated significant reductions on the Hyperactivity and Attention Problem Scales as well as on the more broadly based Internalizing Problems Scale, Externalizing Problems Scale, and Behavioral Symptoms Index. The BASC results confirm that the changes observed in cognitive functioning on the TOVA

were accompanied by significant reductions in AD/HD symptoms and psychopathology more generally as observed in the patients' daily lives.

EFFICACY VERSUS EFFECTIVENESS

Before addressing specific criticisms, it would be instructive to contrast the respective goals and methodology of efficacy and effectiveness studies in clinical research. This comparison places the research methodology of Rossiter and La Vaque (1995) and Rossiter (2003) in perspective.

Kazdin (2003) places efficacy and effectiveness research at the opposite ends of a continuum of research methodologies employed in the assessment of psychotherapy treatment outcomes. Efficacy studies are conducted under laboratory and quasi-laboratory conditions with an emphasis on experimental controls that insure internal validity. The goal of the efficacy study is to determine whether benefits associated with a treatment are due to the treatment itself as opposed to chance or confounding factors such as spontaneous remission, the composition of the patient sample, placebo response, the effects of being assessed, etc. To achieve this goal, the efficacy study employs randomized clinical trials in which patients are recruited and randomly assigned to treatment or control groups (e.g., no-treatment, waiting list, placebo or sham treatment) that serve as a basis for comparison. Efficacy studies employ highly structured treatment protocols that are uniformly applied to all patients and are closely monitored for therapist compliance. Patients are carefully screened to rule out the presence of comorbid disorders and other potentially confounding factors in order to produce homogeneous patient groups. Efficacy studies usually provide only the treatment being evaluated to the experimental group and not the full range of treatments that might ordinarily be available to patients with the disorder being studied.

However, treatments that are demonstrated to be efficacious are not necessarily effective when applied in clinical settings (Weisz, Weiss, & Donenberg, 1992). That is, they may lack external validity. The failure of "efficacious" treatments to successfully transfer to clinical settings may be due, in part, to the very experimental controls that define the efficacy study and insure internal validity (Weisz & Weiss, 1989). For example, Zimmerman, Mattia, and Posternak (2002) applied inclusion and exclusion criteria commonly used in efficacy studies of antidepressant medications to 346 outpatients diagnosed with major depression. Of the original group of patients, only 41 (12%) would have qualified for participation in most efficacy studies of antidepressant medications. Similarly, Leber (1989) notes that the randomized control trial design used in efficacy studies "almost never" (p. S58) enrolls samples of patients that are truly representative of the population of patients with the disorder under study. Leber concludes, "therefore, how well the results of any randomized control trial will predict the performance of a drug in actual medical practice is always a matter of some uncertainty" (p. S58). Although Zimmerman et al. and Leber specifically address pharmacological research, efficacy studies of psychotherapy or biofeedback outcomes are subject to the same limitations regarding external validity.

Effectiveness research, by contrast, is typically conducted in a clinical setting where the usual control procedures that characterize efficacy research are not implemented to the same extent. Effectiveness studies utilize patients who come to a clinic seeking treatment and expecting improvement. Their presenting problems are sufficiently severe to seek

treatment and may include multiple diagnostic categories. The use of heterogeneous patient groups that include comorbid disorders increases the ability to assess the real world effectiveness of the treatment and to generalize from the results (Clarke, 1995). Because the research is conducted in a clinical setting, some compromises in research methodology and experimental controls have to be made for practical and ethical reasons. Treatment may be tailored to meet the needs of the individual patient. The result is that not all patients receive exactly the same treatment. Furthermore, it is the patient, not the clinician, who is ultimately responsible for choosing the treatment. In essence, an effectiveness study can evaluate a treatment as it is actually provided in clinical practice. Effectiveness studies place greater emphasis on external validity than do efficacy studies for which internal validity is of paramount importance. Therefore, effectiveness research has the potential for broad applicability to the real world spectrum of patients as they present for treatment in clinics and hospitals (Clarke, 1995). Because of the less stringent experimental controls, an effectiveness study can demonstrate that a treatment program is clinically effective, but it may not be possible to establish to what extent various elements (e.g., the treatment under study, patient expectations, therapist characteristics, placebo, etc.) contribute to the positive outcomes.

Chambless and Hollon (1998) state that the fundamental question posed by clients presenting for treatment is, "Does the treatment you propose actually work?" (p. 8). However, this is not the question asked by parents seeking treatment for an AD/HD child. Because an effective treatment for AD/HD already exits, they are more likely to ask, "What is the best treatment for my child?" This question cannot be answered by an efficacy study. It requires an effectiveness study that compares neurofeedback and stimulant drugs as they are typically provided to patients in a clinical setting. Both Rossiter and La Vaque (1995) and Rossiter (2003) fall at the effectiveness end of the research continuum and are most appropriately evaluated as such.

CRITICISMS

Control Group

The choice of a control group is central to the design of AD/HD neurofeedback outcome research. As will be demonstrated, the choice is subject to ethical constraints that may not apply to other diagnostic categories. The active treatment control is an infrequently utilized comparison group in psychotherapy outcome research where the use of nontreatment (e.g., waiting list, placebo, sham treatment) controls is the norm.

The Lohr et al. (2001) and Kline et al. (2002) criticism of Rossiter and La Vaque (1995) for failing to include a waiting list control is curious. The waiting list group is the weakest of the nontreatment controls in that it controls for the passage of time and the effects of repeated assessment but little else (Borkovec & Castonguay, 1998; Chambless et al., 1998). With respect to its applicability in an AD/HD effectiveness study, the waiting list control is a particularly poor choice because AD/HD, unlike depression for example, is not a disorder that is self-limiting and prone to spontaneous remission over relatively brief periods of time. Most importantly, a waiting list group provides no control for placebo or nonspecific treatment effects. In addition, the waiting list control is likely to have a higher dropout rate than the active treatment group, thus introducing a potential bias in composition of

the samples and the results (Clarke, 1995). Patients who are experiencing distress that is sufficiently severe and/or chronic to seek treatment are likely to go elsewhere when denied treatment for the 3–12 months needed to complete an AD/HD neurofeedback program.

Lohr et al. (2001) and Kline et al. (2002) contend that the comparison of neurofeedback and standard pharmacological treatment "does not allow conclusions about the efficacy of neurotherapy per se" (pp. 100, 52, respectively). As a caveat, they add that only if the study contains a waiting list control in addition to the two active treatments can unambiguous conclusions be reached about the efficacy of either active treatment. Neither Lohr et al. nor Kline et al. recognize that the active treatment control in Rossiter and La Vaque (1995) is a class of drugs with a 30-year history of placebo-controlled studies. They believe that using a stimulant drug as an active treatment control is comparable to using a behavioral intervention. "When Neurotherapy is extended to psychological disorders, the experimental procedures and criteria of judging the efficacy of neurofeedback are the same as those for the evaluation of other psychological treatments" (Lohr et al., 2001, p. 98). The critics' position reflects a lack of familiarity with the extensive literature in pharmacological research relating to research design and active control studies. The orthodox position in pharmacology research is that a study seeking to establish the efficacy of a new drug requires that the new drug be compared to a placebo to rule out the influence of nonspecific factors. If the purpose is to compare the efficacy of a new drug to a standard drug, a placebo control is still required (Leber, 1989; Temple & Ellenberg, 2000). In a three-armed study of this type, a waiting list control is not an acceptable alternative to a placebo control. In the event that both the experimental and standard drugs were superior to the waiting list group but did not differ significantly from each other, the possibility that nonspecific factors accounted for the improvement in both the experimental and drug groups could not be ruled out. The same logic applies to studies comparing nondrug treatments or nondrug to drug treatments.

The concept of evaluating a new treatment by comparing it to an existing treatment (e.g., best alternative treatment, O'Leary & Borkovec, 1978; minimal treatment control, Weiss & Weisz, 1990; usual care control, Clarke, 1995; standard or routine treatment, Kazdin, 2003) is not unknown in the psychotherapy outcome literature. Clarke (1995) takes the position that this type of control does not adversely affect the internal validity of outcome studies and that it has the greatest ecological validity of all common control conditions. Barkley (1993) recommends comparing EEG biofeedback to an alternative and/or placebo therapy. Chambless and Hollon (1998) conclude that the direct comparisons of competing treatments provide the "most stringent tests of all" (p. 8) and have the added advantage that they control for the nonspecific effects that are common to all treatments.

The overriding objection to the use of any nontreatment control in an AD/HD effectiveness study is that it would be an ethically questionable choice (La Vaque & Rossiter, 2001). The Nuremberg Code published in 1947 and the Declaration of Helsinki (World Medical Association Declaration of Helsinki [WMA], 2000) adopted in 1964 and amended in 1975, 1983, 1989, 1996, and 2000 have served as the core guides of ethical principles for human subject research on a worldwide basis for the past half century. The Article from the Declaration of Helsinki most relevant to Rossiter and La Vaque (1995) and other studies evaluating the effectiveness of neurofeedback for AD/HD is

> Article C.29: The benefits, risks, burdens and effectiveness of a new method should be tested against those of the best current prophylactic, diagnostic, and therapeutic methods. This does not exclude the use of placebo, or no treatment, in studies where no proven prophylactic, diagnostic or therapeutic method exists (p. 3045).

Although two exceptions to the restrictions on the use of placebo controls were adopted by the WMA in 2002, neither applies to studies evaluating the effectiveness of neurofeedback for AD/HD. Article C.29 makes clear that subjects in a study should not be assigned to a nontreatment (waiting list, sham or placebo control) group when a safe and effective treatment is already available.

One of the benefits that may be offered to patients who are assigned to the placebo or nontreatment control group is the opportunity to receive the experimental treatment once it is determined to be safe and effective. From an ethical viewpoint, this is not an acceptable alternative. Offering treatment to the nontreatment control subjects after the study has been completed in 3 (Rossiter, 2003) to 12 (Monastra et al., 2002) months does not meet the requirements of Article C.29, because the proven standard treatment is still withheld during the term of the study.

The efficacy of methylphenidate and dextroamphetamine in treating AD/HD is well established (Barkley, 1990; Brown et al., 1986; MTA Cooperative Group, 1999; Rapoport et al., 1978, 1980; Spencer et al., 1995; Swartwood et al., 1998; Wender, Reimherr, Wood, & Ward, 1985; Zeiner, Bryhn, Bjercke, Truyen, & Strand, 1999). Stimulant drugs are the "treatment of choice" for AD/HD and have been the standard for over 30 years. Therefore, the ethical standards of the Declaration of Helsinki dictate the use of an active treatment control because a nontreatment control would withhold a standard, effective treatment from the control subjects. It should be noted that Article C.29 of the Declaration, although accepted by Japan and the European countries, is quite controversial in the United States and has been explicitly rejected by the American Medical Association.

It is sometimes argued that if a patient randomized to a control condition gives informed consent and the delay in treatment causes minimal harm, no ethical standards have been violated (Temple & Ellenberg, 2000). This argument makes the assumption that the right to waive ethical constraints on the clinician/researcher lies with the patient. This assumption is erroneous. Ethical standards are imposed on clinicians and researchers by state licensing boards and professional organizations. The patient does not have the right to waive ethical constraints. No licensing board or professional organization would accept a properly executed informed consent form as justification for otherwise unethical conduct. Furthermore, as Glantz (1996) points out, the issue of obtaining genuinely informed consent that would allow children to be assigned to a nontreatment control group is ethically troublesome.

Aside from ethical concerns, a number of conditions must be met to justify using an active treatment rather than a nontreatment control (Hwang & Morikawa, 1999). Four conditions applicable to Rossiter and La Vaque (1995) and similar studies using stimulant drugs as an active control are:

1. There must be a proven, effective treatment (standard) for the disorder being treated.
2. Placebo-controlled trials of the standard treatment must have demonstrated sensitivity to drug effects.
3. The standard treatment must have assay sensitivity in the particular trial.
4. The treatment comparison must be fair in terms of dose, regimen, endpoints, and/or pharmacological grounds.

Active-treatment-controlled trials can demonstrate efficacy/effectiveness by showing that a new treatment is as good, i.e., equivalent or noninferior, as a proven, effective standard drug or treatment. The validity of this approach is based on the critical assumption that

the standard treatment has a drug effect in that trial. Sensitivity to drug effects and assay sensitivity are related, but not identical, concepts. Sensitivity to drug effects is demonstrated by a history of well-designed placebo-controlled trials that have consistently shown the standard treatment to be superior to placebo. Assay sensitivity applies to a specific trial and requires that the standard treatment produce a drug effect size of the expected magnitude. There is no question that the stimulant drugs are a proven, effective treatment for AD/HD and have demonstrated sensitivity to drug effects in numerous placebo-controlled studies. Furthermore, assay sensitivity of the stimulant drugs can be demonstrated for Rossiter and La Vaque (1995) and Rossiter (2003). In both studies, the proportion of patients showing significant improvement ($>80\%$) and the effect size (0.8 to >1.0) associated with short-term stimulant drug treatment (MED group) are at or above historical standards (e.g., American Academy of Child and Adolescent Psychiatry [AACAP], 2002). Thus, the comparison of neurofeedback to stimulant drugs in Rossiter and La Vaque (1995) and Rossiter (2003) is fair.

Both Rossiter and La Vaque (1995) and Rossiter (2003) meet the conditions outlined by Hwang and Morikawa (1999) for an active treatment control study. Therefore, a non-inferiority/equivalence trial comparing neurofeedback and stimulant drugs is appropriate.

Assignment of Patients

The critics are correct that random assignment of patients to treatment and control conditions is usually preferable to allowing patients to choose their own treatment. Random assignment with a sufficiently large sample insures that the groups being formed are equivalent in all important respects. In an efficacy study, random assignment is required. However, this is not the case in an effectiveness study that utilizes patients who pay for the services they receive. In studies using random assignment, patients are characteristically recruited volunteers who are not charged for services received. In most instances, this type of study requires external funding. The subjects in Rossiter and La Vaque (1995) were not recruited volunteers. They were primarily minors whose parents were seeking treatment for their AD/HD children, paying for the services received, and expecting improvement in their status. It is unreasonable to suggest that these parents would knowingly and willingly pay to have their children randomly assigned to a nontreatment control group or to a treatment they had not chosen. Moreover in a fee for service setting, random assignment would be unethical (Lubar, 1995).

Outcome Measures

The validity of the TOVA, and to a lessor extent that of other CPTs, is an important issue because the TOVA has been widely used in recent years as an outcome measure in neurofeedback studies of AD/HD and other disorders. Lohr et al. (2001, p. 100) criticized the use of the TOVA and BASC on the grounds that (1) they are "nonstandardized measures with minimal ecological validity"; (2) the TOVA "inconsistently discriminates ADHD children from psychiatric control children"; (3) "the correlation of CPT scores and parent and teacher ratings of AD/HD symptoms is moderate at best"; (4) both the TOVA and the BASC are of questionable validity as measures of treatment efficacy. Waschbusch and Hill (2001) and Kline et al. (2002) expressed similar concerns regarding the TOVA.

The assertion that the TOVA and BASC are "nonstandardized measures" (Lohr et al., 2001, p. 100) is erroneous. In general, a test is standardized if it is based on a systematic sampling of the behavior of interest, provides data on the normative sample, has data on reliability and validity, and is administered and scored according to specific instructions. Examination of the BASC Manual (Reynolds & Kamphaus, 1992) or the TOVA Professional Manual (Leark, Dupuy, Greenberg, Corman, & Kindschi, 1996) leaves no doubt that these instruments meet these minimum criteria.

That is not to say that some aspect of the standardization of the TOVA or the BASC cannot be legitimately criticized. Lohr et al. (2001) could have identified and documented specific deficiencies in the standardization procedures for either test. Their failure to do so with respect to the TOVA is puzzling. Riccio, Reynolds, and Lowe (2001) acknowledge that the manuals and supporting research for the commercially available CPTs "do not meet the current professional standards of technical adequacy for testing" (p. 103). Similar criticisms can be leveled at many neuropsychological tests currently in widespread use, e.g., the Halstead–Reitan Neuropsychological Test Battery. The TOVA's technical shortcomings include normative data derived from samples of convenience and insufficient numbers of adult subjects in some age ranges. These deficiencies do not invalidate the test. However, they do suggest that clinicians exercise caution when applying the normative data to patient populations underrepresented in the standardization sample. In addition, the number of validity and reliability studies available when the TOVA was originally published was limited. However in recent years, there have been additional independent studies that support both the validity and reliability of the TOVA (e.g., Fitzgerald, 2001; Forbes, 1998; Li & Wang, 2000; Llorente et al., 2001; Wada, Yamashita, Matsuishi, Ohtani, & Kato, 2000).

The BASC norms are based on a large national sample that is representative of American children with respect to gender, race and ethnicity, clinical and special education categories, and parent education for the Parent Rating Scale (Reynolds & Kamphaus, 2002). The contention that the BASC lacks ecological validity and is not valid as a treatment outcome measure is inconsistent with the critics' position on parent and teacher behavior rating scales. Waschbusch and Hill (2001) maintain that "standardized parent and teacher behavior rating scales have demonstrated ecological validity for the problems associated with ADHD" (p. 162) and are good measures of treatment response. Similarly, Kline et al. (2002) regard parent and teacher behavior rating scales as part of the contemporary "gold standard" (p. 45) for assessing AD/HD.

Barkley and Edwards (1998) do not share the critics' negative assessment of the BASC. They recommend including a comprehensive child behavior rating scale that assesses the full range of child psychopathology including depression, anxiety, withdrawal, aggression, and conduct disorders in addition to inattentive and hyperactive–impulsive behaviors. Specifically, they recommend using the BASC or the Child Behavior Check List. Barkley and Edwards conclude that both scales have "excellent reliable and valid normative data across a wide age range of children makes their incorporation into the assessment protocol quite convenient and extremely useful" (p. 278).

Numerous studies demonstrating the validity of the BASC with specific clinical groups have appeared since the publication of the test (Reynolds & Kamphaus, 1992). Vaughn, Riccio, Hynd, and Hall (1997) found the Attention Problems scale (Parent BASC) effective in identifying children with AD/HD, primarily inattentive type. Ostrander, Weinfurt,

Yarnold, and August (1998) demonstrated that the Attention Problems scale (Parent BASC) was 97% accurate in differentiating combined or inattentive types of AD/HD from the general population. Furthermore, a cutting score of $T > 59$ correctly identified 88% of their sample of 309 AD/HD children. A high score on the Attention Problems scale is essentially pathognomonic for any of the subtypes of AD/HD. Therefore, use of the BASC may negate the need to include a "narrow band" AD/HD behavior rating scale as part of the diagnostic battery recommended by Barkley and Edwards (1998). Conoley et al. (2001) found that the BASC is very sensitive to behavioral changes in individual children and can be administered repeatedly to assess treatment effects. Thus, the evidence supports the conclusion that BASC is well standardized, has high sensitivity with respect to identifying AD/HD patients, and can be validly used to monitor treatment effects.

The critics rely heavily on Barkley (1991) to support their contention that the TOVA lacks ecological validity. However, Barkley reviews a variety of laboratory or analogue methods for assessing the symptoms of AD/HD and has little to say about computer administered CPTs. More to the point, Barkley (1991) predates most of the published research on the TOVA and the other commercial CPTs and is therefore not directly relevant in evaluating the ecological validity of those instruments.

Barkley's current position on the use of CPTs as a tool for diagnosing AD/HD and monitoring stimulant drug effects is informative. Gordon and Barkley (1998, p. 302) conclude that the CPT paradigm is the "most reliable of psychological tests for discriminating groups of ADHD from normal children," is "sensitive to stimulant drug effects among ADHD children and adolescents," is the "only psychological measure that seems to directly assess the core symptoms of the disorder, namely, impulsiveness and inattention," and is minimally influenced by other cognitive factors. These characteristics may account for the fact that the CPT is the most widely studied format for use in AD/HD evaluations (Gordon & Barkley, 1998).

The assertion that CPT scores have only moderate correlations with parent and teacher behavior ratings is not germane to the use of the TOVA in assessing AD/HD. Although child behavior rating scales are widely used diagnostically and as outcome measures, their validity and utility are limited by their subjective nature and generally low interrater reliability. Parent and teacher behavior ratings reflect not only the behavior of the child, but the situation in which it was observed. In addition, they are influenced by a variety of rater variables including gender, education, intelligence, and emotional status at the time the ratings were made (Barkley, 1990). Achenbach, McConaughy, and Howell (1987), in a comprehensive review of agreement among raters of child behavior problems, found mean correlations of 0.27 between parent and teacher ratings and 0.59 between parent ratings. Because correlations between parent–parent and parent–teacher ratings of children's behavior are no more than "moderate," there is no obvious reason to expect that correlations between CPTs and parent and teacher rating scales would be any more robust.

Riccio, Reynolds, et al. (2001) reviewed the research literature on CPTs focusing on the commercially available versions widely used in clinical practice (TOVA, IVA [Integrated Visual and Auditory CPT], Connor's CPT, and Gordon Diagnostic System). Regarding the diagnostic efficacy of the CPTs, they conclude that (1) CPTs have high levels of sensitivity and specificity for all forms of AD/HD, but only when differentiating between AD/HD and normal individuals; (2) a normal performance on CPTs is useful in ruling out AD/HD and related disorders; and (3) CPTs are useful tools for objectively evaluating and documenting

of the presence of symptoms associated with disorders of self-regulation, particularly impulsivity and attentional problems.

The fact that CPTs are less effective in differentiating between AD/HD and other psychiatric disorders is to be expected. Deficits in attention and/or impulse control are not specific to AD/HD but are associated with many psychiatric disorders and most types of CNS dysfunction (Riccio, Reynolds, et al., 2001). There are no core symptoms that are specific to AD/HD or that distinguish AD/HD from normal childhood behavior except in the degree to which they are present. To establish the AD/HD diagnosis, it is necessary to confirm that the patient manifests not only the age-related cognitive deficits (attention and/or impulse control), but also the behavior patterns that differentiate AD/HD from other psychiatric diagnoses (Kamphaus & Frick, 2002). In clinical practice, CPTs are not used in isolation and are never the sole basis for making a diagnosis of AD/HD. Rossiter (2003) and Rossiter and La Vaque (1995) used the TOVA as part of an assessment battery that included a diagnostic interview, intelligence testing, and parent, teacher, and/or patient behavior rating scales. In some cases, academic achievement testing and/or an adult personality questionnaire (Personality Assessment Inventory) were also used.

With respect to the use of CPTs as outcome measures for medication and other forms of treatment, Riccio, Reynolds, et al. (2001) conclude that CPTs offer a rapid, relatively economical, objective measure of the effect of medications on certain components of attention as well as on processing speed and executive control. Psychostimulants, in particular, have been extensively researched and have resulted in improved CPT performance for both control and clinical populations. Depending on the particular combination of stimulant, dosage, and CPT, studies have consistently shown reduced errors of omission and commission and the majority have demonstrated decreased and less variable response time (e.g., Kavale, 1982; Klorman, Brumaghim, Fitzpatrick, & Borgstedt, 1991; Losier, McGrath, & Klein, 1996; Riccio, Waldrop, Reynolds, & Lowe, 2001). Furthermore, CPTs with their demonstrated sensitivity to medication effects on AD/HD would be equally useful in evaluating alternative treatment outcomes. The sensitivity of the TOVA to the effects of both stimulant drugs and neurofeedback in treating AD/HD have been demonstrated by Fuchs et al. (2003), Monastra et al. (2002), Rossiter (2003), and Rossiter and La Vaque (1995).

Riccio, Reynolds, et al. (2001) also reviewed the relationship of CPTs to other cognitive and behavioral measures. They found that the strongest associations were between CPTs and direct behavioral observations, other measures of attention and executive control, and behavioral ratings of children. “The strong relationship between CPT performance and direct observation suggests that CPTs have moderate to high ecological validity” (Riccio, Reynolds, et al., 2001, p. 156).

Collateral Treatments

The critics fault Rossiter and La Vaque (1995) for including patients who were receiving treatments other than neurofeedback for the EEG group and stimulant medication for the MED group. The critics are correct that the provision of collateral treatments to both treatment groups precludes the possibility of attributing improvement in either group strictly to the primary treatment. Thirteen of the 23 MED patients and 5 of the 23 EEG biofeedback patients received behavior modification and/or educational interventions at home and/or at school. If the provision of collateral interventions favored either group, presumably it was

the MED group. The fact that five of the EEG patients received psychostimulant medication is not relevant to their outcomes. All evaluations with those patients were conducted 2 days after stimulant medications were discontinued. The 48-hr washout period is sufficient to insure that the patients were tested in a medication-free state (DuPaul, Barkley, & Connor, 1998). Moreover, stimulant drugs, even with long-term use, have no residual effects on the symptoms of AD/HD after they have been discontinued (e.g., Monastra et al., 2002; Weiss & Hechtman, 1993). The provision of collateral treatments did not bias the results in favor of the EEG group and likely had minimal effect on the outcomes for either group.

More to the point, it is stated in the Discussion section of Rossiter and La Vaque, "*The study demonstrated that a treatment program with EEG biofeedback as the major component led to significant reduction in both cognitive and behavioral symptoms of AD/HD. . . .*" [italics added]. The provision of collateral services to the EEG biofeedback group is irrelevant to the validity of this conclusion.

Eye Movement Artifact

Kline et al. (2002) raise a technical criticism of the neurofeedback protocols currently in use to treat AD/HD. They suggest that the EEG data used for feedback to the patient are contaminated by artifact, primarily produced by eye movement and eye blinks. This is a potentially serious criticism because, if valid, it suggests that factors other than neurofeedback may be responsible for any reduction in AD/HD symptoms or the symptoms of any disorder being treated with neurofeedback.

Although Kline is identified as an "EEG researcher," (p. 47) it does not appear that he is familiar with the neurofeedback hardware and software, or the protocols commonly used in treating AD/HD. Kline et al. (p. 48) state that, (1) "neurotherapists commonly argue that their techniques activate the frontal lobes"; (2) "eye movement is greatest at the front of the head"; (3) "ocular artifact should be a major concern for neurotherapy proponents, yet it seldom seems to be"; (4) and "competent EEG researchers know how to minimize if not eliminate the effects of eye movements on the EEG records, these techniques are not possible in real-time neurofeedback paradigms." Kline et al. clearly imply that (1) AD/HD treatment protocols use frontal training sites in close proximity to the eyes; (2) neurotherapists are unaware of or are unconcerned about potential sources of artifact; (3) neurofeedback systems make no provision for artifact rejection; and as a result (4) neurofeedback is based on artifact-contaminated EEG data.

These propositions are erroneous. Most AD/HD neurofeedback protocols do not employ frontal electrode placements, but rather place active electrodes on, or adjacent to, the motor strip (e.g., Fuchs et al., 2003; Monastra et al., 2002; Rossiter, 1998, 2002, 2003; Rossiter and La Vaque, 1995). Central electrode placements, because they are well posterior to the eyes, greatly reduce eye movement and eye blink artifact at the recording sights (Lubar & Lubar, 1999). Furthermore, AD/HD neurofeedback protocols typically employ software inhibits that reject EEG data that may be contaminated by ocular and EMG artifact. For example, the Biolex (Ver 2.38) software used by the author to construct treatment protocols for both the Lexicor NRS-1620 and the POD home trainers allows the clinician to program multiple conditions that define parameters related to artifact rejection, reinforcement parameters, etc. Two artifact conditions were used by Rossiter and La Vaque (1995)

and Rossiter (1998, 2002, 2003). When δ (0.5–4.0 Hz) and high β (22.0–30.0 Hz) amplitudes exceed predetermined levels, an artifact inhibit is triggered by the software. Artifact rejection levels are initially set at 2.0 times the baseline amplitude of δ for Movement artifact (eye movement, blinking) and 1.7 times the baseline amplitude of high β for EMG artifact. These settings typically result in data rejection rates of $\leq$5% for Movement and $\leq$10% for EMG artifact. The δ and high β amplitudes usually decrease over time and artifact rejection levels are made more stringent. Rossiter (2002) provides details of EEG changes associated with the successful treatment of an AD/HD adolescent.

When the neurofeedback software detects an artifact condition, no EEG data is accumulated, the patient is informed of the artifact condition(s), and reinforcement to the patient is suspended for the duration of the artifact condition. Ocular and EMG artifacts are observable at central sites but are easily identified and managed by the neurofeedback software.

Statistical Analysis

Rossiter and La Vaque (1995) were criticized, and correctly so, for the use of multiple non-alpha protected *t* tests that increased the likelihood of Type I errors. They made 13 planned comparisons for which significant differences were predicted a priori. When the Holm adjustment (Stevens, 1999) was applied to the 13 planned comparisons, all were significant while maintaining the experiment-wise $\alpha = .05$.

AD/HD neurofeedback studies using an active treatment control require a different statistical analysis than would be used with a nontreatment control group. When a new treatment is compared to a nontreatment control, the goal is to demonstrate that the new treatment is superior to the control group. However, when neurofeedback is compared to a stimulant drug group, there is no expectation that neurofeedback will prove to be superior. The 87% improvement rate of the stimulant medication group in Rossiter and La Vaque (1995) and 84% improvement in Rossiter (2003) effectively preclude demonstrating superiority of the neurofeedback group. The alternative is to demonstrate that neurofeedback produces outcomes that are equivalent, or at least noninferior, to those obtained with stimulant drugs.

The intent of Rossiter and La Vaque (1995) was to establish that the neurofeedback and stimulant medication programs each significantly reduced the symptoms of AD/HD and then to compare the outcomes of the two treatment programs to assess relative effectiveness. When the mean change scores for the four TOVA outcome variables were compared, the differences were not statistically significant (two-tailed *t* test, $df = 22$, $p = .36$–$.98$) indicating that the null hypothesis could not be rejected. On the basis of this finding, the authors concluded that the two treatments produced equivalent results. However, the failure to reject the null hypothesis "does not justify the conclusion that the treatments are exactly equivalent, nor does it necessarily justify the conclusion that the difference between treatments is inconsequential" (Hatch, 1996, p.107).

Rogers, Howard, and Vessey (1993) identify three methods to test the hypothesis that two treatments are clinically equivalent: the confidence interval approach (Westlake, 1981); the nonequivalence null hypothesis approach (Anderson & Hauck, 1983); and Bayesian methods (Selwyn, Dempster, & Hall, 1981). The first two methods have the advantage that they require fewer arbitrary decisions (Westlake, 1988).

To demonstrate the equivalence of two treatments, it is first necessary to define an equivalence interval. Because the difference between the competing treatments is likely to be greater than zero, an equivalence interval is defined that is small enough that any difference between the two treatments falling within the interval can be considered clinically unimportant. In other words, group means or proportions are considered equivalent if they differ by less than some small amount that has no clinical significance (Hatch, 1996). However, the decision as to what constitutes a meaningful difference between the two treatment outcomes in any given case may not be a straightforward matter. Where there is no recognized clinical standard that can be used to define the equivalence interval, an alternative would be to set a minimum percentage by which the two groups must differ if they are to be considered clinically equivalent. Hwang and Morikawa (1999) recommend that the equivalence interval should be one half or less of the historical effect size of the active control drug. For the stimulant drugs, the historical effect size is 0.8–1.0 (e.g., AACAP, 2002; Vitiello, 2001). The United States Food and Drug Administration (FDA) takes the position that two formulations of a drug can be considered bioequivalent if they differ by 20% or less on certain bioavailability parameters (Hatch, 1996). In cases where there are no generally agreed upon clinical benchmarks that can be used to define an equivalence interval for a standard treatment or drug, a difference of 20% or less between the standard treatment and the new treatment is the common de facto standard (Schuirmann, 1987).

However, an equivalence or noninferiority study using an active treatment control requires a much larger sample than a placebo-controlled superiority study. Hwang and Morikawa (1999) estimate that the sample size necessary to demonstrate equivalence in an active controlled study is at least four times that for a placebo-controlled superiority trial, when the equivalence margin is one half or less of the control drug effect size. The significant increase in the required sample size is the primary disadvantage of the active-treatment-controlled equivalence study.

One purpose of Rossiter (2003) is to test the general hypothesis that the EEG biofeedback program leads to patient outcomes that are equivalent, or at least noninferior, to stimulant drugs treatment for AD/HD. The proportion of patients demonstrating clinically significant improvement on the TOVA was chosen as the subject for analysis. The following formulas are adapted from Hatch (1996) and Hwang and Morikawa (1999) to test this hypothesis. Let EEG = test treatment, MED = standard active control treatment, δ = noninferiority/equivalence margin, d = the difference in the proportions ($p_{\text{eeg}} - p_{\text{med}}$) of the two groups, and s_d = sample standard error of d. The noninferiority/equivalence margin chosen was 20% of the proportion of MED patients improved (Schuirmann, 1987).

After the equivalence interval has been determined, the nonequivalence null hypothesis approach leads to the following two 1-tailed null hypotheses:

$$\mathrm{H}_{01}: d \leq \delta_1 \quad \text{and} \quad \mathrm{H}_{02}: d \geq \delta_2.$$

The alternative hypothesis, H_a, assumes that the difference between the proportions fall within the equivalence interval:

$$\mathrm{H}_a: \delta_1 < d < \delta_2.$$

The test statistic for the noninferiority trial is

$$z_1 = (d + \delta)/s_d.$$

The null hypothesis can be rejected and it can be concluded that EEG is noninferior (but not equivalent) to MED if $z_1 > z_{1-\alpha}$ (a one-sided α-level test). For testing equivalence, the test statistics for two 1-sided α-level tests are:

$$z_1 = (d + \delta)/s_d \quad \text{and} \quad z_2 = (\delta - d)/s_d.$$

EEG and MED are equivalent if

$$z_1 \quad \text{and} \quad z_2 > z_{1-\alpha}.$$

Theoretically, both z tests must be done in order to reject both null hypotheses. However in practice, only the test having the larger p value needs to be carried out. Because the two tests are completely dependent, if the larger p value is rejected then the test having the smaller p value will be rejected as well (Hatch, 1996; Rogers et al., 1993). Because only a single test is performed, the Type I error rate is equal to the nominal α-level and no adjustment is required for multiple tests (Hatch, 1996; Rogers et al., 1993).

As an alternative to testing null hypotheses, a confidence interval approach can be used to test for noninferiority and equivalence. The equivalency confidence interval (CI) is expressed at the $1 - 2\alpha$ level of certainty rather than the more customary $1 - \alpha$ level. The $\text{CI}_{90\%}$ is computed using the following equation:

$$\text{CI}_{90\%} = d + z_\alpha(s_d).$$

Equivalence of EEG and MED is demonstrated if the confidence interval is contained within the equivalence interval.

CONCLUSIONS

Lohr et al. (2001), Waschbusch and Hill (2001), and Kline et al. (2002) criticized Rossiter and La Vaque (1995) for (1) using an active treatment control rather than a waiting list or other nontreatment control; (2) forming treatment groups by matching patients from intact groups rather than through random assignment; (3) providing collateral treatments to a number of patients in each group; (4) using assessment instruments that were not standardized, lacked ecological validity, could not discriminate between AD/HD and other psychiatric disorders, and were not sensitive to treatment effects; (5) providing artifact contaminated EEG feedback; and (6) conducting multiple non-alpha protected t tests. It was necessary to address these criticisms because some could be directed at Rossiter (2003), Fuchs et al. (2003), and Monastra et al. (2002), and other studies evaluating the effectiveness of neurofeedback for AD/HD, particularly those using stimulants as an active treatment control.

The critics are correct in faulting Rossiter and La Vaque (1995) for using multiple non-alpha protected t tests. However, when the Holm adjustment (Stevens, 1999) was applied to the 13 planned comparisons, all were significant while maintaining the experiment-wise $\alpha = .05$. The other criticisms of Rossiter and La Vaque are either not valid or are not supported by the research literature as presented by the authors. They are based on the critics' unsupported opinions (assessment instruments); rely on redefining the study as efficacy rather than effectiveness study (provision of collateral treatments, nonrandom

assignment of patients); or reflect a lack of familiarity with the relevant research literature (active treatment control) and neurofeedback equipment and AD/HD treatment protocols (artifact contaminated EEG biofeedback).

Therefore, Fuchs et al. (2003), Monastra et al. (2002), Rossiter (2003), and Rossiter and La Vaque (1995) do provide evidence that neurofeedback is an effective treatment for AD/HD. Collectively, they warrant the conclusion that patient outcomes obtained with neurofeedback are comparable (equivalent, noninferior, not significantly different) to those obtained with stimulant drug therapy.

However, there are broader issues to be considered. Specifically, what research methodology is appropriate for studies evaluating the effectiveness of neurofeedback for AD/HD and other disorders and who should make that determination? It is the author's opinion that the uncritical acceptance and implementation of research models developed for psychotherapy, pharmacology, and/or medical research is premature and ill-advised. Neurofeedback researchers need to develop models that are appropriate for neurofeedback outcome studies. This may require a bottom-up approach that involves a synthesis of elements from other disciplines, as appropriate, and the development of new or modified design elements to address unique aspects of the EEG, neurofeedback, and the patient populations being studied. To the extent that their research methodology differs from the currently accepted models, neurofeedback researchers need to clearly explain their rationale and be prepared to defend their choices.

REFERENCES

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101,* 213–232.

American Academy of Child and Adolescent Psychiatry. (2002). Practice parameter for the use of stimulant medications in the treatment of children, adolescents, and adults. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*(2S), 26S–49S.

Anderson, S., & Hauck, W. W. (1983). A new procedure for testing equivalence in comparative bioavailability and other clinical trials. *Communications in Statistics—Theory and Methods, 12,* 2663–2692.

Barkley, R. A. (1977). A review of stimulant drug research with hyperactive children. *Journal of Child Psychology and Psychiatry, 18,* 137–165.

Barkley, R. A. (1990). *Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment.* New York: Guilford Press.

Barkley, R. A. (1991). The ecological validity of laboratory and analogue assessment methods of ADHD symptoms. *Journal of Abnormal Child Psychology, 19,* 149–178.

Barkley, R. A. (1992). Is EEG biofeedback treatment effective for ADHD children? *Ch.A.D.D.er Box,* pp. 5–11.

Barkley, R. A. (1993). Continuing concerns about EEG biofeedback/neurofeedback. *ADHD Report, 1*(3), 1–3.

Barkley, R. A., & Edwards, G. (1998). Diagnostic Interview, Behavior Rating Scale, and the Medical Examination. In R. A. Barkley (Ed.), *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment* (2nd ed., pp. 263–293). New York: Guilford Press.

Borkovec, T. D., & Castonguay, L. G. (1998). What is the meaning of empirically supported therapy? *Journal of Consulting and Clinical Psychology, 66*(1), 136–142.

Brown, R. T., Wynne, M. E., Borden, K. A., Clingerman, S. R., Geniesse, R., & Spunt, A. L. (1986). Methylphenidate and cognitive therapy in children with attention deficit disorder. A double-blind trial. *Developmental and Behavioral Pediatrics, 7,* 163–172.

Chambless, D. L., Baker, M. J., Baucaom, D. H., Beutler, L. E., Calhoun, K. S., Crits-Christoph, P., et al. (1998). Update on empirically validated therapies II. *The Clinical Psychologist, 51*(1), 3–16.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*(1), 7–18.

Clarke, G. N. (1995). Improving the transition from basic efficacy research to effectiveness studies: Methodological issues and procedures. *Journal of Consulting and Clinical Psychology, 63,* 718–725.

Conoley, C. W., Graham, J. M., Neu, T., Craig, M. C., O'Pry, A., Cardin, S. A., et al. (2001, August). *The efficacy of solution focused family therapy with four aggressive and oppositional acting children.* Paper presented at the annual meeting of the American Psychological Association, San Francisco.

DuPaul, G. J., Barkley, R. A., & Connor, D. E. (1998). Stimulants. In R. A. Barkley (Ed.), *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment* (2nd ed., pp. 263–293). New York: Guilford Press.

Fitzgerald, R. L. (2001). Statistical reliability of the test of variables of attention. *Dissertation Abstracts International, 61*(7–B), 3895.

Forbes, G. B. (1998). Clinical utility of the Test of Variables of Attention (TOVA) in the diagnosis of attention deficit hyperactivity disorder. *Journal of Clinical Psychology, 54,* 461–476.

Fuchs, T., Birbaumer, N., Lutzenberger, W., Gruzelier, J. H., & Kaiser J. (2003). Neurofeedback treatment for attention-deficit/hyperactivty disorder in children: A comparison with methylphenidate. *Applied Psychophysiology and Biofeedback, 28*(1), 1–12.

Glantz, L. H. (1996). Conducting research with children: Legal and ethical issues. *Journal of the American Academy of Child and Adolescent Psychiatry, 35*(10), 1283–1291.

Gordon, M., & Barkley, R. A. (1998). Tests and observational measures. In R. A. Barkley (Ed.), *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment* (2nd ed., pp. 263–293). New York: Guilford Press.

Hatch, J. P. (1996). Using statistical equivalence testing in clinical biofeedback research. *Biofeedback and Self-Regulation, 21*(2), 105–119.

Hwang, I. K., & Morikawa, T. (1999). Design issues in noninferiority/equivalence trials. *Drug Information Journal, 33,* 1205–1218.

Kamphaus, R. W., & Frick, P. J. (2002). *Clinical assessment of child and and adolescent personality and behavior* (2nd ed.). Needham Heights, MA: Allyn & Bacon.

Kavale, K. (1982). The efficacy of stimulant drug treatment for hyperactivity: A meta-analysis. *Journal of Learning Disabilities, 15,* 280–289.

Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston: Allyn & Bacon.

Kline, J. P., Brann, C. N., & Loney, B. R. (2002). A cacophony in the brainwaves: A critical appraisal of neurotherapy for attention-deficit disorders. *The Scientific Review of Mental Health Practice, 1,* 44–54.

Klorman, R., Brumaghim, J. T., Fitzpatrick, P. A., & Borgstedt, A. D. (1991). Methylphenidate speeds evaluation processes of attention deficit disorder adolescents during a continuous performance test. *Journal of Abnormal Child Psychology, 19,* 263–283.

La Vaque, T. J., & Rossiter, T. R. (2001). The ethical use of placebo controls in clinical research: The Declaration of Helsinki. *Applied Psychophysiology and Biofeedback, 26*(1), 23–37.

Leark, R. A., Dupuy, T. R., Greenberg, L. M., Corman, C. L., & Kindschi, C. L. (1996). *Test of Variables of Attention Professional manual version 7.0.* Los Alamitos, CA: Universal Attention Disorders.

Leber, P. D. (1989). Hazards of inference: The active control investigation. *Epilepsia, 30*(Suppl. 1), S57–S63.

Li, X., & Wang, Y. (2000). A preliminary application of the Test of Variables of Attention (TOVA) in China. *Chinese Mental Health Journal, 14*(3), 149–152.

Llorente, A. M., Amado, A. J., Voight, R. G., Berretta, M. C., Fraley, J. K., Jensen, C. L., et al. (2001). Internal consistency, temporal stability, and reproducibility of individual index scores of the Test of Variables of attention in children with attention-deficit/hyperactivity disorder. *Archives of Clinical Neuropsychology, 16*(6), 535–546.

Lohr, J., Meunier, S., Parker, L., & Kline, J. P. (2001). Neurotherapy does not qualify as an emprically supported behavioral treatment for psychological disorders. *The Behavior Therapist, 24*(5), 97–104.

Losier, B. J., McGrath, P. J., & Klein, R. M. (1996). Error patterns on the continuous performance test in non-medicated and medicated samples of children with and without ADHD: A meta-analytic review. *Journal of Child Psychology and Psychiatry, 37,* 971–987.

Lubar, J., & Lubar, J. (1999). Neurofeedback assessment and treatment for attention deficit/hyperactivity disorders. In J. R. Evans & A. Abarbanel (Eds.), *Introduction to quantitative EEG and neurofeedback* (pp. 103–143). San Diego, CA: Academic Press.

Lubar, J. F. (1995). Neurofeedback for the management of attention-deficit/hyperactivity disorder. In M. S. Schwartz & Associates (Eds.), *Biofeedback: A practitioner's guide* (2nd ed., pp. 493–522). New York: Guilford Press.

Monastra, V. J., Monastra, D. M., & George, S. (2002). The effects of stimulant therapy, EEG biofeedback, and parenting style on the primary symptoms of attention-deficit/hyperactivity disorder. *Applied Psychophysiology and Biofeedback, 27*(4), 231–249.

MTA Cooperative Group. (1999). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Archives of General Psychiatry, 56,* 1073–1086.

O'Leary, K. D., & Borkovec, T. D. (1978). Conceptual, methodological, and ethical problems of placebo groups in psychotherapy research. *American Psychologist, 33,* 821–830.

Ostrander, R., Weinfurt, K. P., Yarnold, P. R., & August, G. J. (1998). Diagnosing attention deficit disorders with the Behavior Asssessment System for Children and the Child Behavior Checklist: Test and construct validity analyses using optimal discriminant classification trees. *Journal of Consulting and Clinical Psychology, 66,* 660–672.

Rapoport, J. L., Buchsbaum, M. S., Weingartner, H., Zahn, T., Ludlow, C., & Mikkelsen, E. J. (1980). Dextroamphetamine: Its cognitive and behavioral effects in normal and hyperactive boys and normal men. *Archives of General Psychiatry, 37,* 933–943.

Rapoport, J. L., Bucksbaum, M. S., Zahn, T. P., Weingartner, H., Ludow, C., & Mikkelsen, E. J. (1978). Dextroamphetamine: Cognitive and behavioral effects in normal prepuberal boys. *Science, 199,* 560–563.

Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior Assessment System for Children manual*. Circle Pines, MN: American Guidance Service.

Reynolds, C. R., & Kamphaus, R. W. (2002). *The clinician's guide to the Behavior Assessment System for Children*. New York: Guilford Press.

Riccio, C. A., Reynolds, C. R., & Lowe, P. A. (2001). *Clinical applications of continuous performance tests: Measuring attention and impulsive responding in children and adults*. New York: Wiley.

Riccio, C. A., Waldrop, J. J., Reynolds, C. R., & Lowe, P. A. (2001). Effects of stimulants on the continuous performance test (CPT): Implications for CPT use and interpretation. *Journal of Neuropsychiatry and Clinical Neurosciences, 13,* 326–335.

Rogers, J. L., Howard, K. I., & Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychological Bulletin, 113*(3), 553–565.

Rossiter, T. R. (1998). Patient directed neurofeedback for AD/HD.*Journal of Neurotherapy, 2,* 54–64.

Rossiter, T. R. (2002). Neurofeedback for AD/HD: A ratio feedback case study and tutorial. *Journal of Neurotherapy, 6*(3), 9–35.

Rossiter, T. R. (2003). The effectiveness of neurofeedback and stimulant drugs in treating AD/HD: Part II. Replication. Manuscript submitted for publication.

Rossiter, T. R., & La Vaque, T. J. (1995). A comparison of EEG biofeedback and psychostimulants in treating attention deficit hyperactivity disorders. *Journal of Neurotherapy, 1,* 48–59.

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics, 15*(6), 657–680.

Selwyn, M. R., Dempster, A. P., & Hall, N. R. (1981). A Bayesian approach to bioequivalence for the 2×2 changeover design. *Biometrics, 37,* 11–21.

Spencer, T., Wilens, T., Biederman, J., Faraone, S. V., Ablon, J. S., & Lapey, K. (1995). A double-blind, crossover comparison of methylphenidate and placebo in adults with childhood-onset attention-deficit hyperactivity disorder. *Archives of General Psychiatry, 52,* 434–443.

Stevens, J. S. (1999). *Intermediate statistics: A modern approach* (2nd ed.). Mahwah, NJ: Erlbaum.

Swartwood, M. O., Swartwood, J. N., Lubar, J. F., Timmermann, D. L., Zimmerman, A. W., & Muenchen, R. A. (1998). Methylphenidate effects on EEG, behavior, and performance in boys with ADHD. *Pediatric Neurology, 18*(3), 244–250.

Temple, R. T., & Ellenberg, S. S. (2000). Placebo-controlled trials and active-controlled trials in the evaluation of new treatments. *Annals of Internal Medicine, 133*(6), 455–463.

Vaughn, M., Riccio, C., Hynd, G., & Hall, J. (1997). Diagnosing ADHD (predominantly inattentive and combined subtypes): Discriminant validity of the Behavior Assessment System for Children (BASC) and the Achenbach Parent and Teacher Rating Scales. *Journal of Clinical Child Psychology, 26*(4), 349–357.

Vitiello, B. (2001). Methylphenidate in the treatment of children with attention-deficit hyperactivity disorder. *Canadian Medical Association Journal, 165*(11), 1505–1506.

Wada, N. Y., Yamashita, Y., Matsuishi, T., Ohtani, Y., & Kato, H. (2000). The Test of Variables of Attention (TOVA) is useful in the diagnosis of Japanese male children with attention deficit hyperactivity disorder. *Brain and Development, 22*(6), 378–382.

Waschbusch, D. A., & Hill, G. P. (2001). Alternative treatments for children with Attention-Deficit/Hyperactivity Disorder: What does the research say? *The Behavior Therapist, 24*(8), 161–171.

Weiss, G., & Hechtman, L. T. (1993). *Hyperactive children grown up: ADHD in children, adolescents, and adults* (2nd ed.). New York: Guilford Press.

Weiss, B., & Weisz, J. R. (1990). The impact of methodological factors on child psychotherapy outcome research: A meta-analysis for researchers. *Journal of Abnormal Child Psychology, 18,* 639–670.

Weisz, J. R., & Weiss, B. (1989). Assessing the effects of clinic-based psychotherapy with children. *Journal of Consulting and Clinical Psychology, 57,* 741–746.

Weisz, J. R., Weiss, B., & Donenberg, G. R. (1992). The lab versus the clinic: Effects of child and adolescent psychotherapy. *American Psychologist, 47,* 1578–1585.

Wender, P. H., Reimherr, F. W., Wood, D., & Ward, M. (1985). A controlled study of methylphenidate in the treatment of attention deficit disorder, residual type, in adults. *American Journal of Psychiatry, 142,* 547–552.

Westlake, W. J. (1981). Bioequivalence testing—A need to rethink. *Biometrics, 37,* 591–593.
Westlake, W. J. (1988). Bioavailability and bioequivalence of pharmaceutical formulations. In K. E. Peace (Ed.), *Biopharmaceutical statistics for drug development* (pp. 329–352). New York: Marcel Decker.
World Medical Association Declaration of Helsinki. (2000). Ethical principles for medical research involving human subjects. *JAMA, 284*(23), 3043–3045.
Zeiner, P., Bryhn, G., Bjercke, C., Truyen, K., & Strand, G. (1999). Response to methylphenidate in boys with attention-deficit hyperactivity disorder. *Acta Paediatrica Scandinavia, 88*(3), 298–303.
Zimmerman, M., Mattia, J. I., & Posternak, M. A. (2002). Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *American Journal of Psychiatry, 159*(3), 469–473.