

Letter to the Editor

Correction Concerning Howard et al. (2005)

J. Howard, C. Sparkman, H. Cohen, G. Green, and H. Stanislaw

We would like to take this opportunity to correct some misrepresentations of our study “A comparison of intensive behavior analytic and eclectic treatments for young children with autism” (Howard et al., 2005) that have appeared in this journal and elsewhere.

Although we welcome thoughtful and informed criticisms of our research, we think the field and the best interests of children with autism are seriously jeopardized by the dissemination of unfounded statements and insinuations about the study and those involved in it. To attenuate those harms, we respectfully offer the following pertinent facts about the Howard et al. (2005) study:

- The nonpublic agency that provided intensive behavior analytic treatment to the experimental group in our study has never refused to serve a child with autism who was referred to its in-home or center-based programs from any of its 17 different funding sources (Special Education Local Planning Agencies, Regional Centers, and school districts), as long as there were adequate resources to staff and supervise the programming.
- Children who participated in our study were actually served and referred to the study by two different Regional Centers.
- Neither the staff of the nonpublic agency nor any Regional Center staff person ever directed a child to be placed in any particular intervention program.
- All 61 IEP/IFSP teams responsible for the educational placements of the children in the study acted in accordance with state and federal laws governing special education placement and procedures. That is, those teams -- not the authors of the study or any other single individuals -- decided which type of intervention children would receive.
- Children were not randomly assigned to groups because of serious ethical and practical constraints; additionally, doing so would have violated the aforementioned laws. Rather, participants were selected for our study from larger pools of children in the three types of programs in such a way as to constitute matched treatment and comparison groups. Consistent with recommended tactics for field research of this nature, the groups were matched as closely as possible on variables that have been found to correlate with treatment outcomes in previous studies. We did not merely assume that the matching procedures resulted in groups that were substantially similar when the study began; rather, we demonstrated it empirically, as described in our paper. The small number of variables on which the groups differed significantly pre-treatment were controlled for in our statistical analyses, also described in our paper.
- Our study, which was published in a respected peer-reviewed journal, substantially replicated the methods and results of a previous study conducted by independent investigators in another country (Eikeseth, Smith, Jahr, & Eldevik, 2002). It also corroborates several other published studies of early intensive behavior analytic treatment for autism.

For interested readers, the Howard et al. (2005) study is available at www.sciencedirect.com. The full reference is Howard, J., Sparkman, C., Cohen, H., Green, G., & Stanislaw, H. (2005). A comparison

of intensive behavior analytic and eclectic treatments for young children with autism. Research in Developmental Disabilities, 26, 359-383.

As we have in the past, we encourage anyone who has questions or concerns about our study to contact us directly. We will make every effort to respond.

Author Contact Information

Jane S. Howard, PhD, BCBA
Psychology Department
California State University, Stanislaus
801 W. Monte Vista Avenue
Turlock, CA 95382
jhoward@asherah.csustan.edu

Coleen Sparkman, MA, CCC-SLP
The Kendall School
1115 14th St
Modesto, CA 95354
(209) 572-0102

Howard Cohen, PhD
Valley Mountain Regional Center
P.O. Box 692290
Stockton, CA 95269-2290
Phone: (209) 473-0951

Gina Green, PhD, BCBA
San Diego State University
San Diego, California
Phone: 619-518-4990

Harold Stanislaw, PhD
Department of Psychology
California State University, Stanislaus
801 W. Monte Vista Avenue
Turlock, CA 95382
Phone: (209) 667-3386
e-mail: hstanislaw@csustan.edu

Letter To The Editor

Contributions of Howard and Colleagues (2005) to Early Intensive Behavioral Treatment

Tristram Smith

Howard, Sparkman, Cohen, Green, and Stanislaw (2005) made two breakthroughs: They and others forged a community-wide collaboration to give families of newly diagnosed children with autism information about a variety of interventions, including early intensive behavioral treatment (EIBT) and other intensive services, and to offer families a choice among these interventions, which are provided at no cost to families (Cohen, 2003; Cohen, Amerine-Dickens, & Smith, 2006; Region 6 Autism Connection, 2004). With this collaboration in place, they were able to complete one of the first studies comparing EIBT to other services of equal intensity, finding that EIBT was associated with much more favorable outcomes (Howard et al., 2005).

However, commenting on Howard et al.'s (2005) study, Schoneberger (2006) cites a sometime-collaborator of theirs (me) as a detractor of their research methods (pp. 209, 211-212). To evaluate this assertion, it is informative to look at my remarks to Schoneberger in their entirety—a brief email I sent in response to a request from him for my thoughts on the study:

Dear Mr. Schoneberger,

I think the paper is a very useful contribution to the literature. Limitations of the study include non-random assignment to groups, limited information about the interventions, and unclear procedures for rating the presence or absence of symptoms of autism. However, the groups appear fairly comparable prior to treatment, and the comprehensive assessments and use of independent examiners are strengths. Also, the results are impressive.

I hope these comments are helpful.

Tristram Smith

About my email, Schoneberger says only, “Tristram Smith (T. Smith, personal communication, July 25, 2005) has reported that one of the Howard et al. study's ‘limitations’ is its use of nonrandom assignment” (p. 209), that “according to Tristram Smith (T. Smith, personal communication, July 25, 2005), another weakness of the Howard et al. study is that it provides ‘limited information about the interventions’” (p. 212), and that I did not identify problems with the assessment protocol as another limitation (p. 211). Obviously, this characterization of our correspondence is quite one-sided and misleading.

To redress this imbalance, I would underscore several points: Although Schoneberger (2006) enumerates the potential problems with non-random assignment, he does not mention that a design with matched groups (as in the Howard et al., 2005, study) is generally considered to be the next best approach (Cook & Campbell, 1979). Howard et al.'s use of this design (2005) is justified because random assignment was not possible. (Treatment was funded by public agencies with a mandate to provide free and appropriate services to participants, and therefore unable to assign them randomly to groups).

Moreover, the design addresses many of the potential problems that Schoneberger (2006) identified. For example, because the groups in Howard et al.'s (2005) study did not differ significantly on pretreatment measures of IQ, language, and adaptive skills, the superior outcome of the EIBT groups relative to the comparison groups cannot be explained by pretreatment differences between groups in children's level of functioning. Although "limited information about the interventions" is another concern, it pertains more to the write-up of the study than to the study design itself and might have been resolved simply by requesting additional information from the authors. Finally, although Schoneberger (2006) is correct that a few participants in the Howard et al. (2005) study did not receive the measures that were part of the standard assessment protocol, it is also true that the large majority received did complete this protocol (ranging from 42 to 54 of the 57 participants across measures). While it obviously would have been preferable if the same protocol had been administered to all participants, this seems unlikely to explain why the EIBT group obtained higher test scores than the comparison groups after treatment.

Howard et al.'s (2005) study was the first comparison-controlled study of an EIBT model other than the UCLA approach developed by Lovaas and colleagues (Lovaas & Smith, 2003). It also was one of the first EIBT studies to include multiple outcome measures obtained by independent examiners, yet it found consistent advantages for EIBT over eclectic services across all measures. Remarkably, many EIBT children raised their standardized test scores into the average range after only 14 months of intervention (compared to 2-3 years of intervention in previous studies).

Howard et al. (2005) acknowledged the limitations of their study and the need for replications (pp. 378-379), but, given the many strengths that their study also had, they appropriately concluded that the results provide evidence for the efficacy of EIBT. Their innovative community-wide collaboration, comparison of EIBT with other services of equal intensity, and favorable EIBT outcomes stand as significant achievements.

References

- Cohen, H. G. (2003). Pyramid building: Partnership as an alternative to litigation. In O. I. Lovaas (Ed.), *Teaching Individuals with Developmental Delays: Basic intervention techniques* (pp. 375-386). Austin, TX: PRO-ED.
- Cohen, H. Amerine-Dickens, M., & Smith, T. (2006). Early intensive behavioral treatment: Replication of the UCLA Model in a community setting. *Journal of Developmental and Behavioral Pediatrics*, 27, S145-S155.
- Cook, T.D. and Campbell, D.T. (1979). *Quasi-Experimentation: Design and analysis for field settings*. Rand McNally, Chicago, Illinois.
- Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities*, 26, 359-383.
- Lovaas, O. I., & Smith, T. (2003). Early and intensive behavioral intervention in autism. In A. E. Kazdin & J. Weisz (Eds.), *Evidence-based psychotherapies for children and youth* (pp. 325-340). New York: Guilford.
- Region 6 Autism Connection (2004). *Early Intensive Behavioral Treatment 4-way Agreement*. Stockton, CA: Author.

Schoneberger, T. (2006). EIBT research after Lovaas (1987): A tale of two studies. *Journal of Speech-Language Pathology and Applied Behavior Analysis*, 1, 210-217.

Author Information

Tristram Smith, Ph.D.
Strong Center for Developmental Disabilities
Golisano Children's Hospital at Strong
University of Rochester School of Medicine and Dentistry
Box 671
601 Elmwood Avenue
Rochester, NY 14642
Phone: (585)273-3515
e-mail: Tristram_Smith@URMC.Rochester.edu

**ADVERTISING IN THE JOURNAL OF SPEECH - LANGUAGE PATHOLOGY AND
APPLIED BEHAVIOR ANALYSIS**

The prices for advertising in one issue are as follows:

1/4 Page: \$50.00 1/2 Page: \$100.00 Full Page: \$200.00

**If you wish to run the same ad in multiple issues for the year, you are eligible for the
following discount:**

**1/4 Pg.: \$40 - per issue
1/2 Pg.: \$75 - per issue
Full Page: \$150.00-per issue**

An additional one-time layout/composition fee of \$25.00 is applicable

For more information, or to place an ad, contact Halina Dziewolska by phone at

(215) 462-6737 or e-mail at: halinadz@hotmail.com

EIBT Research after Lovaas (1987): A Tale of Two Studies

Ted Schoneberger

Abstract

Since the publication of Lovaas's (1987) seminal paper, serious questions have surfaced regarding design features that compromise the validity of treatment efficacy data resulting from studies of early intensive behavioral treatment (EIBT) for children with autism. Lovaas and his colleagues have acknowledged the legitimacy of some of these questions, and guidelines have emerged to improve the validity of future efficacy studies: (1) use random assignment of participants; (2) use uniform assessment protocols; (3) document enough methodological detail to support replication. Two recent studies are examined in reference to their compliance with these guidelines (Howard, Sparkman, Cohen, Green, & Stanislaw, 2005; Sallows and Graupner, 2005). Findings indicate that different levels of compliance result in different degrees of threats to internal validity.

Keywords: autism, autism spectrum disorders, early intensive behavioral treatment, Lovaas, random assignment, uniform assessment, replication.

Introduction

Since the publication of Lovaas's (1987) seminal study, a growing body of research has been conducted to document the treatment efficacy of early intensive behavioral treatment (EIBT) for children with autism spectrum disorders (e.g., Birnbrauer & Leach, 1993; Harris, Handleman, Gordon, Kristoff, & Fuentes, 1991; McEachin, Smith, & Lovaas, 1993; Scheinkopf & Siegal, 1998). However, serious questions have been raised about the validity of much of this research. Lovaas's (1987) study in particular, which was designed to examine the efficacy of treatment offered at the UCLA Young Autism Project, has been subjected to considerable criticism (e.g., Gresham & MacMillan, 1997; Schopler, Short, & Mesibov, 1989). While much of this criticism was challenged by Lovaas and his colleagues, some of it was acknowledged by them to be valid. Moreover, these acknowledgements served as the basis for emphasizing three research guidelines (among others) that could improve the validity of follow-up treatment efficacy studies. Unfortunately, the application of these guidelines has not always been consistent.

The current paper summarizes those criticisms of the Lovaas (1987) study that Lovaas and his colleagues have acknowledged as valid, and it summarizes the research guidelines that evolved from these criticisms. Further, two recently published EIBT studies are reviewed and evaluated relative to their compliance with these guidelines. The first study was conducted by Howard, Sparkman, Cohen, Green, and Stanislaw (2005) and the second was conducted by Sallows and Graupner (2005).

Criticisms of the Lovaas (1987) Research

Criticisms of the Lovaas (1987) research have been addressed by Tristram Smith, a colleague of Lovaas's and Research Director of the Multi-Site Young Autism Project. Specifically, Smith (in Smith, Groen, & Wynn, 2000) noted two criticisms with which Lovaas and his associates reportedly concur:

First, assignment to groups was based on whether or not therapists were available to provide intensive treatment rather than on a more arbitrary procedure, such as the use of a random numbers table. Thus, *assignment could have been biased* [italics added]. Second,

because children were referred to outside examiners, they received a variety of different tests rather than a uniform assessment protocol. Hence, *assessment results may have been unreliable* [italics added]. (p. 270)

While concurring with these criticisms, this concurrence was nonetheless qualified. For example, Lovaas and his colleagues have reportedly been dubious about the importance of the second criticism. Nevertheless, to address these (and other) criticisms relative to future research, they emphasized “the need for replication to confirm the results” (Smith et al., 2000, p. 270). Also inherent in their responses are three guidelines to be addressed by follow-up treatment efficacy studies: (1) random assignment of participants to treatment conditions; (2) use of uniform assessment protocols across all participants; and (3) documentation of sufficient methodological detail to allow for independent replication. Unfortunately, mixed results are evident in published follow-up studies relative to these recommendations. In the next section, we examine two prominent, recently-published EIBT studies to illustrate this point.

The Treatment Efficacy Study of Howard, Sparkman, Cohen, Green, and Stanislaw (2005)

Howard et al. (2005) studied 61 children diagnosed with either autistic disorder or pervasive developmental disorder-not otherwise specified (PDD-NOS). These participants were referred by nonprofit agencies (“regional centers”) whose primary function is to meet the case management needs of people with developmental disabilities. To be eligible for the study, participants had to satisfy the following criteria: (a) receive a diagnosis of Autism or PDD-NOS before the 4th birthday; (b) be exposed to English as the primary language at home; (c) be available to begin treatment before the age of four years; and (d) have not received more than 100 hours of treatment prior to participating in the study.

Treatment Conditions

The children in the Howard et al.'s (2005) study participated in one of three different, multi-component treatment conditions. Each is summarized below:

Group #1: Intensive Behavior Analytic Intervention (IBT). In the IBT group, participants under 3 years of age received 1:1 intervention for 25 to 30 hours per week, while those over 3 years of age received 1:1 intervention for 35 to 40 hours per week. The IBT participants received treatment across multiple settings including school and home. Using discrete trial training, incidental teaching, as well as “other behavior analytic procedures” (Howard et al., p. 7), 50 to 100 trials per hour were presented. Further, parents were provided with training in fundamental behavior analytic strategies, maintenance and generalization data collection techniques, and methods for implementing their children’s treatment programs “outside of regularly scheduled intervention hours” (p. 7). Parents were also required to attend meetings with agency staffers one to two times per month.

Group #2: Autism Educational Programming (AP). The AP participants received 25 to 30 hours per week of 1:1 or 1:2 interventions delivered in public school classrooms designated to serve students with autism. These participants received a range of interventions including activities derived from the TEACCH model, discrete trial training, Picture Exchange Communication System (PECS) training, and sensory integration therapy.

Group #3: Generic Educational Programming (GP). The GP participants received 15 hours per week of 1:6 interventions in special education, preschool classrooms designated to serve either early intervention or communicatively handicapped students. *Developmentally appropriate* instructional activities were employed, emphasizing “exposure to language, play activities, and a variety of sensory experiences” (Howard et al., p. 8). In addition, a certified speech-language pathologist provided most of

the participants with language therapy once or twice a week.

Results

Standardized assessments targeting cognitive, nonverbal, receptive and expressive language, and adaptive skills were administered to the participants during intake and at follow-up (after approximately 14 months of treatment). At intake all three groups had “similar” (Howard et al., p. 11) mean scores on all but one measure. The only difference achieving statistical significance was in the domain of nonverbal skills. Moreover, for all three groups, the mean standard scores across most skill domains were considerably below 100. At follow-up, the differences in the mean scores of the participants in the AP and GP groups were *not* statistically significant. On the other hand, “the IBT group had higher mean scores in all domains than the other two groups combined; and those differences were statistically significant” (Howard et al., p. 11).

At follow-up, the IBT group's mean standard scores for the cognitive, nonverbal, communication, and motor skills domains were within normal range; the only domain in which the AP and GP groups scored in the normal range was motor skills. Thirteen IBT group participants exhibited gains in their IQ scores “from one standard deviation or more below average (i.e., IQ of 85 or lower) at intake to within one standard deviation of average or above (i.e., IQ of 86 or higher) at follow-up” (Howard et al., p. 11). Three other IBT participants whose intake IQ scores were in the normal range (i.e., 84, 89, and 97) exhibited follow-up gains (i.e., from 84 to 122, 89 to 114, and 97 to 102). At intake none of the AP participants exhibited IQ scores within the normal range; at follow-up, two exhibited IQs within the normal range. The IQ scores of 3 participants in the GP group changed from one (or more) standard deviations below average (at intake) to within normal range (at follow-up). Finally, two GP participants who exhibited intake IQ scores within the normal range displayed a *decrease* in their IQ scores at follow-up.

Compliance Guideline #1: Random Assignment

The Howard et al. study did not follow the first guideline. Specifically, a quasi-experimental pretest-posttest, nonequivalent groups design was employed. Participants were assigned to the groups by their respective individual education plan (IEP) or individual family service plan (IFSP) teams where “parental preferences weighed heavily” (p. 6). More specifically, each child's team considered “a range of educational options” which included (but was not limited to) placement in one of the three groups.

The Howard et al. study is appropriately characterized as a *nonequivalent* group design because the participants were not randomly assigned to the three conditions. McGuigan (1997) has defined *random assignment* as “a procedure that assures that each member of a population or universe has an equal probability of being selected” (p. 89). According to Durso and Mellgren (1989), random assignment is the “most important” method of controlling extraneous variables, and “the prerequisite for a *true experiment*” (p. 106). Similarly, Graziano and Raulin (2004) have described random assignment as “the most basic and single most important control procedure” (p. 207). Failure to randomly assign is a “basic weakness” (Kerlinger, 1973, p. 321) of nonequivalent group designs. Similarly, Tristram Smith (T. Smith, personal communication, July 25, 2005) has reported that one of the Howard et al. study's “limitations” is its use of nonrandom assignment. Thus, by failing to use random assignment, the design employed by Howard et al. is not considered a true experiment, but rather quasi-experimental (Cozby, 2001).

Howard et al. acknowledge that their use of nonrandom assignment constitutes a limitation of the

study. However, they also assert that the three groups were “very similar” on “key” pretreatment, dependent measures, and that this is the “main purpose of random assignment” (p. 15). But *is* this its main purpose? According to Kerlinger (1973), random assignment is used to provide the rationale for assuming that groups are equal “in *all* characteristics” (p. 127; emphasis added), not just equal (or, worse yet, merely “similar”) with respect to “key” (Howard et al., p.15) or “pertinent” (Kerlinger, p. 321) dependent variables. When “randomization is not used . . . it is *not possible* [italics added] to assume that the groups are equal” (p. 322). For McGuigan (1997), “the great value of randomization is that it randomly distributes extraneous effects, whatever they may be, over the experimental and control conditions” (p. 90). When “we do *not* randomly assign participants to groups, . . . we can expect confounds” (McGuigan, 1997, p. 90). Thus, lacking the presumption of equivalence, “we must consider the likelihood that alternative hypotheses may account for the results.” (McBurney, 1998, p. 249). Two such alternative hypotheses shall now be considered.

Alternative Hypothesis #1: In accounting for the results reported by Howard et al., alternative hypothesis #1 centers on how differentially motivated to help their children the participants' parents were across the three conditions. Remember that, according to Howard et al., the parental preferences regarding educational placement (i.e., regarding assignment to a particular treatment condition) “weighed heavily” (p. 6). Further, recall that in the IBT treatment condition “parents received training in basic behavior analytic strategies, assisted in the collection of maintenance and generalization data, implemented programs with their children outside of regularly scheduled intervention hours, and met with the agency staff 1-2 times a month” (p. 7). In the other two treatment conditions, no such comparable demands on the parents were identified. Thus, one plausible alternative hypothesis is that those parents who were willing to actively participate in their children's treatment were more motivated to help them change, and thus more likely to choose the IBT condition, while those less motivated were more likely to opt for the AP and GP conditions.

Graziano and Raulin (2004) offered a similar alternative hypothesis in their discussion of a hypothetical pretest-posttest nonequivalent group design conducted to determine “whether eliminating food containing the additives thought to increase hyperactivity will help hyperactive children” (pp. 224-225). In this hypothetical quasi-experiment, the researcher formed the groups by asking the parents whether or not they were willing to effectuate for their children a 4 week diet which excluded the food additives:

The children of those parents who were willing to expend the effort were put in the experimental group, and those who were not were put in the control group. The serious confounding in this procedure is that the experimental and the control groups are different in terms of parents' willingness to try the dietary restrictions. In other words, the dietary restriction treatment is confounded with parents' willingness to cooperate. We might assume that parents who are willing to do everything necessary to change their child's diet in hope of decreasing their child's hyperactivity may be more motivated to help their child to change. Any posttreatment differences between the groups on measures of hyperactivity might be due to either factor: dietary restriction or parental willingness to cooperate. (p. 225)

Similarly, in Howard et al.'s study, type of autism treatment (IBT, AP, GP) is confounded with parent's willingness to actively participate in treatment. Note that it is *not* the differing roles played by parents across the treatment conditions that is considered a confound; rather, it is differences in parental motivation that is the confound. Post-treatment differences between the IBT group and the two other conditions might be due to the IBT parents being more motivated than the parents in the other conditions to assist their children in changing. There are a number of plausible explanations why this increased

motivation may serve as a confound. Perhaps the IBT parents, being more motivated, implemented the treatment program at times, and in locales, that exceeded what was required of them. Or perhaps the IBT parents were more motivated because their children had been more responsive to their earlier attempts to teach them (i.e., *prior* to their entry into the study), thus giving the parents some hope of success when eventually exposed to a highly structured training regimen.

Alternative hypothesis #2. This hypothesis concerns bias associated with the influence of the non-parental members of the IEP/IFSP teams, including the special education and case management professionals. As fiduciaries within a federally mandated special education process, these team members had both a legal and ethical responsibility to advocate assigning to the IBT group only those children for whom such a placement would be “appropriate” based on the child’s IEP/IFSP. An “appropriate” placement is one which permits the child to “benefit educationally” from the instruction (Bateman & Linden, 1992/1998, pp. 143-144). So, although the parents’ choice of treatment conditions may have, according to Howard et al., “weighed heavily,” the other members of the team doubtless used their expertise to influence the eventual decision.

Indeed, during part of the time in which the Howard et al. study was conducted, representatives of Therapeutic Pathways (the service provider under whose auspices the Howard et al study was conducted) were empowered to play a determinative role in the IEP/IFSP process, *regardless of parental preferences*. Specifically, the Howard et al study was conducted “from 1996 through 2003” (Howard et al., p. 6). In their 1999 manual “In-home Programs for Young Children with Autism,” Therapeutic Pathways required the parents to grant Therapeutic Pathways the ultimate decision making power regarding the range and content of the treatment program, as well as eventual school placement¹. A reasonable assumption is that Therapeutic Pathways staffers used this power to place in the IBT condition those participants who, in their professional judgment, were more likely to benefit from the program, and to refer the other participants to the other conditions. If this is the case, then experimenter-based, biased assignment to groups provides an obvious alternative explanation of the results.

Consider also another example of the decisive role played by Therapeutic Pathways. As a “nonpublic” (i.e., private) agency, the representatives of the service provider were free to refuse to treat any child who, in their judgment, would not benefit from their program. Howard et al. do not inform us whether or not the service provider exercised this option and, if they did, the criteria that were used. Obviously, if they refused to treat some children whom they judged would not benefit from the program, then this, too, suggests bias.

Compliance with Guideline 2: Use of Uniform Assessment Protocol

Although not identified by Smith (T. Smith, personal communication, July 25, 2005) as a limitation of the study, a close reading of the published paper (Howard et al., pp. 8-10) indicates that the researchers failed to use a uniform assessment protocol during intake and follow-up, thus risking unreliable measurement. Consider these assessment issues in detail across these domains: Cognitive skills, nonverbal skills, receptive and expressive language, and adaptive skills.

Assessment of Cognitive skills. The participants’ cognitive skills were assessed using a number of different instruments at intake. Specifically, 42 participants were assessed using the *Bayley Scales of Infant Development-Revised*; 10 participants were assessed using the *Wechsler Primary Preschool Scales of Intelligence-Revised*; 3 were assessed using the *Developmental Profile-II*; and 2 were assessed using the Stanford-Binet Intelligence Scale. Three additional instruments were used to assess each of three children, respectively: *Differential Abilities Scale*, *Developmental Assessment of Young Children*, and *Psychoeducational Profile Revised*.

At follow-up, “the test used . . . varied with the chronological ages of the child” (p. 9). A majority of the participants (i.e. 47 children) participated in cognitive assessments using an instrument that was *not* used at intake (i.e., the *Wechsler Primary Preschool Scales of Intelligence-Revised*). However, three instruments used during intake assessment of cognitive skills were also used at follow-up, albeit for a small number of participants. Specifically, 4 participants were assessed at follow-up using the *Bayley Scales of Infant Development*; 3 were assessed using the *Stanford-Binet Intelligence Scale*; and 2 were assessed using the *Differential Abilities Scale*.

Assessment of Nonverbal skills. At intake, the nonverbal skills of 48 participants were assessed using the *Merrill-Palmer Scale of Mental Tests*; and one participant was assessed using the *Stanford-Binet Performance Test*. At follow-up, 54 participants were assessed with the same instrument used during intake (i.e., the *Merrill-Palmer Scale of Mental Tests*). One participant was assessed at follow-up using the *Leiter International Performance Scale Revised*, which was not used at intake.

Assessment of Receptive and Expressive Language. The *Reynell Developmental Language Scales* were used to assess the receptive and expressive language skills of 46 participants at intake. Other instruments used at intake included the Rossetti Infant-Toddler Language Scale (for 5 participants); the *Receptive-Expressive Emergent Language Scales-Revised* (for 3 participants), and the *Preschool Language Scale* (for 3 participants). One child was assessed with three instruments at intake, including the *Toddler Developmental Assessment*, the *Peabody Picture Vocabulary Test-3rd Edition*, the *Expressive Vocabulary Test*, and the *Developmental Profile-II* language scale. In the case of one client, there was no assessment of receptive and expressive language skills.

At follow-up, 47 participants were assessed using the same instrument that was used during intake (the *Reynell Developmental Language Scales*). Other instruments used at follow-up (only some of which had also been used during intake) included the *Sequenced Inventory of Communication Development-Revised Edition* (for 3 participants); *Peabody Picture Vocabulary Test-3rd edition* along with the *Expressive Vocabulary Test* (for 2 participants); *Preschool Language Scales-3* (for 2 participants); and the *Expressive One-Word Picture Vocabulary Test* along with the *Receptive One-Word Picture Vocabulary Test* (for 1 participant). In the case of 6 participants, the receptive and expressive language skills were not measured at all during follow-up.

Assessment of Adaptive Skills. At intake, 54 participants were assessed using the *Vineland Adaptive Behavior Scales*. Other instruments used were: the personal adjustment or self-help scales of the *Denver Developmental Screening Test II* (3 participants), *Developmental Profile-II* (1 participant), and the *Rockford Infant Development Evaluation Scales* (1 participant). Two participants were not assessed at all during intake. At follow-up, 56 participants were assessed using the *Vineland Adaptive Behavior Scales*, and 6 participants received no assessment.

Compliance with Guideline #3: Replicability

According to Tristram Smith (T. Smith, personal communication, July 25, 2005), another weakness of the Howard et al. study is that it provides “limited information about the interventions”. This weakness makes it next to impossible to replicate. Similarly, Smith's third, remaining criticism (i.e., that there were “unclear procedures for rating the presence or absence of symptoms of autism”) also implies that it would be difficult to attempt replication without greater clarity in this area (not to mention that this weakness raises the issue of possible biased sampling and biased assignment to groups).

Other Methodological Problems

According to Howard et al., follow-up assessments were conducted by examiners who were *not* blind to treatment condition assignments of the participants. This raises the obvious issue of examiner bias in favor of IBT, which is clearly a threat to internal validity. Howard et al. argue, however, that given the substantial number of different examiners reportedly used, it is “just as likely” (p. 15) that some assessors were biased *against* IBT as *for* IBT. Unfortunately, no evidence is offered to support this assertion.

Indeed, although the examiners are identified as “independent” of both the investigators and the treatment programs, it remains unclear how these follow-up examiners were funded and thus, how independent they actually were. As Howard et al. reported, one of the agencies funding the research was Valley Mountain Regional Center (VMRC). During the time frame in which the Howard et al. study was conducted, VMRC also contracted with independent vendors to do assessments (N. McGonigle, personal communication, June 14, 2006). Were any of these VMRC-funded vendors used to conduct any of the follow-up assessments in the Howard et al. study? If so, then at least some of these follow-up assessments were funded by the *same* agency that provided funding for the research (i.e., VMRC). This raises the issue of conflict of interest, thereby strengthening concerns about potential (presumably unintentional) bias on the part of the examiners.

An additional potential conflict of interest problem concerns Howard et al.'s third author, an individual who served as the Clinical Director of VMRC during the study's time frame. As the supervisor of some VMRC staffers involved in making placement decisions, what role, if any, did he play in influencing their decisions? There is clearly a conflict between (1) his role as a fiduciary with a responsibility to see to it that only those who are likely to benefit from the IBT treatment package are so assigned and (2) his role as researcher with a responsibility to avoid biased assignment to groups. How was this conflict in roles resolved? Howard et al. do not say.

The Treatment Efficacy Study of Sallows and Graupner (2005)

Sallows and Graupner (2005) studied 24 children with a diagnosis of autism. At intake these children met six criteria: (1) they ranged in age between 24 to 42 months; (2) they had a Mental Development Index “ratio estimate” (i.e., MA divided by CA) of 35 or more; (3) they were “neurologically within normal limits.” (Note, however, that children with abnormal EEGs or controlled seizures were accepted as determined by a pediatric neurologist, and no child was excluded based on this criterion.); (4) a developmental diagnosis was established by “independent child psychiatrists” (p. 420); and (5) the diagnosis met the DSM-IV and Autism Diagnostic Interview-Revised criteria for autism. (A “trained examiner” administered both instruments.) In addition, (6) “there were no parental criteria for involvement beyond agreeing to the conditions in the informed consent document” (p. 420).

Treatment Conditions

Each participant was assigned to one of two groups. The nature of each group is summarized below:

Group #1: Clinic-Directed. This group received treatment “replicating the parameters of the UCLA intensive behavioral treatment” (Sallows & Graupner, p. 420). Specifically, the group received “the treatment procedure and curriculum . . . initially described by Lovaas (Lovaas et al., 1981) except that no aversives were used.” Additional procedures, buttressed by subsequent research (e.g., Koegel & Koegel, 1995) were also employed (Sallows & Graupner, p. 422). During the first two years, participants received an average of 38 hours per week of direct treatment. Thereafter, as the children began school, the

weekly direct treatment hours were gradually decreased. This group “received 6 to 10 hours per week of in-home supervision from a senior therapist and weekly consultation by the senior author or clinic supervisor” (p. 421). Further, “parents were instructed to attend weekly team meetings and were encouraged to extend the impact of treatment by practicing the newly learned material with their child throughout the day” (p. 420).

Group #2: Parent-Directed. This group received essentially the same treatment as Group #1, except that it was less intense. Specifically, “parents in the parent-directed group chose the number of weekly treatment hours provided by therapists” (p. 421). Thus, during the first two years, participants averaged 31.5 hours per week of direct treatment “with the exception that one family chose to have 14 hours both years” (p. 421). As with Group #1, direct treatment hours were then slowly decreased as the child entered school. Further, this group “received 6 hours per month of in-home supervision from a senior therapist (typically a 3-hour session every other week) and consultation every 2 months by the senior author or clinic supervisor” (p. 421). As with Group #1, parents were told to attend weekly team meetings and urged to practice their newly acquired skills throughout the day with their children.

Results

Sallows and Graupner (2005, p. 417) reported that the “outcome after 4 years of treatment, including cognitive, language, adaptive, social, and academic measures, was similar for both groups.” For example, on average, the full scale IQ for all participants showed a 25 point increase. Specifically, the authors noted that

Parent-directed children, who received 6 hours per month of supervision . . . did about as well as clinic-directed children, although they received much less supervision. This was unexpected, and it may have been due in part to parent-directed parents taking on the senior therapist role, filling cancelled shifts themselves, actively targeting generalization, and pursuing teachers and neighbors to find peers for daily play dates with their children. Although many parent-directed parents initially made decisions regarding treatment that resulted in their children progressing slowly . . . , many parents then sought input from treatment supervisors and rapidly learned to avoid making the same mistake twice, becoming quite skillful after a few months. (p. 433)

Compliance with *Guideline #1: Random Assignment*

Sallows and Graupner's study is a product of the Wisconsin Young Autism Project. As participants in the Lovaas' Multi-Site Young Autism Project, these researchers “worked in collaboration with and observed the guidelines set by the National Institutes of Mental Health” (p. 419). Thus, in adherence to NIMH-approved research protocol, preschoolers diagnosed with autism were matched “on pretreatment IQ (Bayler MA divided by CA)” and then “randomly assigned by a UCLA statistician” to the clinic-directed group or the parent-directed group. In short, matched *random* assignment was used, thus satisfying the first guideline. Indeed, it is noteworthy that, while parents clearly had the option to drop out of the study if unhappy with their child's group assignment, “none dropped out upon learning of their group assignment, minimizing bias in selection of participants and group composition” (p. 420).

By randomly assigning participants to groups, this study avoided many of the problems associated with nonrandom assignment (see previous discussion). Further, in employing *matched* random assignment, the study achieved additional benefits. Random assignment is employed to make it more likely that “the difference between subjects that might affect the outcome of the experiment will be even, or averaged out” (Durso & Mellgren, 1989, p. 159). However, by chance, subjects with characteristics

likely to strengthen post-treatment performance may still be disproportionately assigned to one condition over the other. Matching is recommended as a means of addressing this threat to internal validity under certain conditions. Specifically, whenever possible, the researcher should use matching if “there is a subject characteristic that is *highly correlated with the dependent variable*” (Durso & Mellgren, 1989, p. 162). Citing a number of studies (e.g., Bibby, Eikeseth, S., Martin, N. T., Mudford, O. C., & Reeves, D., 2002; Lovaas, 1987), Sallows and Graupner (2005) identified IQ as one of the “most commonly noted predictors” (p. 419) of post-treatment outcome. So, participants in the Sallows & Graupner study were first matched on pretreatment IQ and then randomly assigned to either the clinic-directed or parent-directed group, thereby bolstering the benefits achieved when only simple random assignment is used.

Compliance with Guideline #2: Uniform Assessment

During intake, pretreatment measures were taken of all participants, using five different instruments: (a) the *Bayley Scales of Infant Development*, Second Edition; (b) the *Merrill-Palmer Scale of Mental Tests*; (c) *Reynell Developmental Language Scales*; (d) *Vineland Adaptive Behavior Scales* and (e) the *Early Learning Measure* (an experimental assessment tool developed by Smith, Buch, & Gamby, 2000). In addition, direct observation, reports of other professionals, and parent interviews were used to determine the developmental history, “supplemental treatments” history, and presence/absence of functional speech. In sum, a uniform assessment protocol appears to have been followed at intake, thus adhering to the second guideline. However, at follow-up, which occurred annually over a four-year period, this guideline was not strictly followed. With the exception of the Early Learning Measure (which was only given a second time, after several months of treatment), all of the pretreatment instruments were administered at follow-up to at least *some* of the participants. Instruments administered *only at* follow-up (i.e., not at intake) included the *Wechsler Preschool and Primary Scale of Intelligence-Revised*; the *Wechsler Intelligence Scale for Children--WISC-III*; the *Leiter R*; the *Clinical Evaluation of Language Fundamentals*, Third Edition; the *Woodcock Johnson III Tests of Achievement*; the *Personality Inventory for Children*; and the *Child Behavior Checklist*. Explaining the use of these other instruments, Sallows and Graupner reported that “as children grew older or became too advanced for the norms of pretreatment tests, we used other age-appropriate tests” (p. 421).

Compliance with Guideline #3: Replicability

The guideline calling for sufficient information to allow for replication was largely satisfied. The researchers not only employed a widely available, detailed set of instructional strategies (e.g., Lovaas et al., 1981), they also provided additional detail within the body of their paper. For example, they reported that no aversives were used, and that in addition to Lovaas et al. (1981), other treatment procedures inspired by more recent research (e.g., Koegel & Koegel, 1995) were also used. Additional treatment strategies used included conducting only two to three training trials at a time, and using continuous, immediate, and powerful reinforcement. “Between these brief (initially 30 seconds long) learning periods, staff members played with the children to keep the process more like play than work, generalize learned material into more natural settings, and continue to build social responsiveness” (see Sallows & Graupner, 2005, p.422-423 for a discussion of treatment strategies used).

Conclusion

Almost two decades have elapsed since the publication of Lovaas's (1987) original treatment efficacy study. During the first decade, many of the criticisms of that study appeared in print, along with subsequent responses to these criticisms by Lovaas and his colleagues. These responses acknowledged the

legitimacy of some of the criticisms, thereby suggesting some minimal, necessary guidelines to be followed by follow-up research. In this paper, two recent treatment efficacy studies were described and evaluated in reference to the three guidelines. Results indicated that when these guidelines are followed, the results are better equipped to withstand criticisms; and when they are not followed, the results are more vulnerable to these critiques.

Utilizing a pretest-posttest nonequivalent groups design, the Howard, Sparkman, Cohen, Green, and Stanislaw (2005) study failed to demonstrate the superiority of early intensive behavioral treatment over that provided by special day classes in public schools. This failure was further exacerbated by Howard et al.'s use of a non-uniform assessment protocol (suggesting unreliability of measurement), as well as their failure to provide anything approaching adequate information about the details of each treatment condition, thus making replication impossible. In comparison, the Sallows and Graupner (2005) study utilized matched, random assignment, an assessment protocol more closely approximating uniformity, and sufficient detail to allow for replication, thus advancing our knowledge of the efficacy of parent-directed early intensive behavioral treatment (EIBT) as a less intrusive, less costly alternative to clinic-directed EIBT treatment.

References

- Bateman, B. D., & Linden, M. A. (1992/1998). *Better IEPs: How to develop legally correct and educationally useful programs* (3rd ed.). Longmont, CO: Sopris West.
- Bibby, P., Eikeseth, S., Martin, N. T., Mudford, O. C., & Reeves, D. (2002). Progress and outcome for children with autism receiving parent-managed intensive interventions. *Research in Developmental Disabilities, 23*, 81-104.
- Birnbrauer, J. S., & Leach, D. J. (1993). The Murdoch Early Intervention Program after two years. *Behaviour Change, 10*, 63-74.
- Cozby, P. C. (2001). *Methods in behavioral research* (7th ed.). Mountain View, CA: Mayfield.
- Durso, F. T., & Mellgren, R. L. (1989). *Thinking about research: Methods and tactics of the behavioral scientist*. St. Paul, MN: West Publishing.
- Graziano, A. M., & Raulin, M. L. (2004). *Research methods: A process of inquiry* (5th ed.). Boston, MA: Pearson.
- Gresham, F. M., & MacMillan, D. L. (1997). Autism recovery? An analysis and critique of the empirical evidence on the Early Intervention Project. *Behavioral Disorders, 22*, 185-201.
- Harris, S., Handleman, J., Gordon, R., Kristoff, B., & Fuentes, F. (1991). Changes in cognitive and language functioning of preschool children with autism. *Journal of Autism and Developmental Disabilities, 21*, 281-290.
- Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities, 26*, 359-383.
- Kerlinger, F. N. (1973). *Foundations of behavioral research* (2nd ed.). NY: Holt, Rinehart and Winston.
- Koegel, R. L., & Koegel, L. K. (1995). *Teaching children with autism: Strategies for initiating positive*

interactions and improving learning opportunities. Baltimore: Brookes.

- Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*, 55, 3-9.
- Lovaas, O. I., Ackerman, A. B., Alexander, D., Firestone, P., Perkins, J., & Young, D. (1981). *Teaching developmentally disabled children: The me book*. Austin, TX: Pro-Ed.
- Lovaas, O. I., Smith, T., & McEachin, J. J. (1989). Clarifying comments on the young autism study: Reply to Schopler, Short, and Mesibov. *Journal of Consulting and Clinical Psychology*, 57, 165-167.
- McBurney, D. H. (1998). *Research methods* (4th ed.). Pacific Grove, CA: Brooks/Cole.
- McEachin, J. J., Smith, T., & Lovaas, O. I. (1993). Long-term outcome of children with autism who received early intensive behavioral treatment. *American Journal of Mental Retardation*, 43, 589-595.
- McGuigan, F. J. (1997). *Experimental psychology: Methods of research* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Sallows, G. O., & Graupner, T. D. (2005). Intensive behavioral treatment for children with autism: Four-year outcome and predictors. *American Journal of Mental Retardation*, 110, 417-438.
- Smith, T., Groen, A. D., & Wynn, J. W. (2000). Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal of Mental Retardation*, 105, 269-285.
- Smith, T., & Lovaas, O. I. (1997). The UCLA Young Autism Project: A reply to Gresham and MacMillan. *Behavioral Disorders*, 22, 202-218.
- Smith, T., McEachin, J. J., & Lovaas, O. I. (1993). Comments on replication and evaluation of outcome. *American Journal on Mental Retardation*, 97, 385-391.
- Therapeutic Pathways (June 1999, version 4). *In-home programs for young children with autism* manual.

Notes

- 1 . While I am not a lawyer, surrendering broad, decision making power to one member of an IEP/IFSP team seems inconsistent with federal law. Specifically, I doubt that parents can voluntarily surrender their rights, or the rights of other IEP/IFSP members, given that all are considered equal partners in the IEP/IFSP process (see Bateman, B. D. & Linden, M. A., 1992/1998).

Author's Note: *Disclaimer: The views expressed here are not necessarily shared by my employer, the Stanislaus County Office of Education (Modesto, CA), nor by my fellow employees.*

Author Contact Information:

Ted Schoneberger, P.O. Box 157, Turlock, CA 95381, Phone: (209) 556-5655, E-mail: TSberger@aol.com