American Journal on Mental Retardation, 1997, Vol. 102, No. 3, 238–249

Intensive Behavioral Treatment for Preschoolers With Severe Mental Retardation and Pervasive Developmental Disorder

Tristram Smith
Washington State University

Svein Eikeseth
University of Kansas

Morten Klevstrand
University of Oslo

O. Ivar Lovaas
University of California, Los Angeles

From archival records, we assessed outcomes achieved by preschoolers with both severe mental retardation and autistic features: (a) an experimental group (n = 11), which received intensive behavioral treatment, and (b) a comparison group (n = 10), which received minimal treatment. At intake (mean CA = 3.08 years), the groups did not differ significantly on any variable. At follow-up children in the experimental group obtained a higher mean IQ and evinced more expressive speech than did those in the comparison group. Behavior problems diminished in both groups. Results indicate that intensively treated children achieved clinically meaningful gains relative to the comparison group but remained quite delayed.

Building on research that began in the early 1960s, investigators in the mid-1980s began reporting that many children with autism who received behavioral treatment during their preschool years benefited greatly (cf. Simeonnson, Olley, & Rosenthal, 1987). In the most detailed of these reports, Lovaas and his colleagues (Lovaas, 1987; McEachin, Smith, & Lovaas, 1993) studied the effects of a behavioral treatment program that consisted of 40 hours per week of one-to-one instruction provided in children's homes and communities for 2 or more years. A group of 19 preschool children with autism who received this intensive treatment made major increases in intellectual, academic, adaptive, and socioemotional functioning. By the age of 7 years, 9 of these 19 children (47%) scored in the average range on standardized tests of intelligence and performed satisfactorily in regular first-grade classes. By contrast, in two control groups, consisting of 40 similar children with autism who received minimal treatment, only one child (2.5%) achieved such a favorable outcome. Moreover, the 19 intensively treated children maintained their gains at a follow-up conducted several years after the end of treatment, when

the children averaged 13 years of age. A number of preliminary studies also have yielded encouraging outcomes (Anderson, Avery, DiPietro, Edwards, & Christian, 1987; Birnbrauer & Leach, 1993; Fenske, Zalenski, Krantz, & McClanahan, 1985; Harris, Handleman, Gordon, Kristoff, & Fuentes, 1991; Hoyson, Jamieson, & Strain, 1984). Efforts are underway to replicate these findings with improved research methodologies (Smith, McEachin, & Lovaas, 1993).

The behavioral treatment approaches used by Lovaas and other investigators were not intended exclusively for preschoolers with autism. Rather, they were developed for a wide range of children with delays (Lovaas & Smith, 1988). Therefore, in addition to conducting replications focusing on children with autism, we have been interested in examining the extent to which behavioral treatment would result in favorable outcomes for other populations.

One population that merits investigation is preschoolers who have severe mental retardation as well as autistic behaviors. Such children are typically underrepresented in treatment outcome research, despite their obvious need for professional attention, because of a diagnostic problem: It is difficult to determine which children should receive a diagnosis of both autism and mental retardation and which should receive a diagnosis of mental retardation alone. This difficulty arises because many (perhaps most) children with severe mental retardation display high levels of autistic behaviors (Wing, 1981). Consequently, investigators have been uncertain when to assign a diagnosis of autism along with severe mental retardation.

To ensure that a diagnosis of autism is warranted, investigators such as Lord and Schopler (1988) have recommended excluding children with severe mental retardation. Consistent with this recommendation, Lovaas and his colleagues (Lovaas, 1987; McEachin et al., 1993) applied an IQ cutoff, requiring children to achieve a ratio IQ of 37 or higher to participate in the study. Other investigators have accepted children at all levels of intellectual functioning but have reported on very few children with severe mental retardation (Anderson et al., 1987; Birnbrauer & Leach, 1993; Fenske et al., 1985; Harris et al., 1991; Hoyson et al., 1984).

At present, therefore, virtually no information exists about the outcomes achieved with behavioral treatment for preschoolers who have both severe mental retardation and autistic features (cf. Guralnick & Bricker, 1987). As a step toward obtaining such information, we evaluated the effectiveness of one form of behavioral treatment, the model developed by Lovaas and his associates, when administered to this population. The evaluation consisted of an analysis of archival data from children seen at the UCLA Young Autism Project and by sites replicating this project.

Method

Participants

All 21 referrals (19 boys, 2 girls) who met the following criteria were included in the present study: (a) chronological age (CA) of 46 months or less at the time of referral; (b) ratio IQ of less than 35, as measured by the Mental Development Index of the Bayley Scales of Infant Development (Bayley, 1969); (c) diagnosis by a licensed psychologist or psychiatrist who was not involved in this research and who had expertise in assessing young children with developmental disabilities, based on *Diagnostic and Statistical Manual* (*DSM-III*) criteria (American Psychiatric Association, 1980), of both mental retardation and pervasive developmental disorder; (d) residence within a one-hour drive of the treatment site in order to allow the provision of services in children's homes and communities; and (e) absence of any major medical limitations.

For the purposes of this study, a *major medical limitation* was defined as any one of the following: (a) a condition requiring prosthetic devices (e.g., cerebral palsy or deafness), (b) an illness that would prevent a subject from receiving treatment for 6 consecutive weeks or more (e.g., metastasized cancer or end-stage renal disease), (c) a syndrome known to have a degenerative course (e.g., Rett disorder), or (d) a diagnosed hereditary disorder (e.g., fragile X syndrome, phenylketonuria, or Down syndrome). Children with these conditions were excluded because the treatment sites did not employ personnel qualified to treat them, such as physicians and physical therapists.

Design

All referrals who met the selection criteria for participants were assigned to one of two groups: (a) an experimental group of 11 boys who received intensive treatment (30 hours or more of one-to-one treatment per week for two or more years), and (b) a comparison group of 8 boys and 2 girls, who received minimal treatment (10 hours per week or less of one-to-one treatment for up to 2 years). Assignment to groups was based on therapist availability: If sufficient therapists were available, participants entered the experimental group and received intensive treatment; otherwise, they entered the comparison group and received minimal treatment.

Six of the 11 experimental participants were enrolled in the UCLA Young Autism Project (Lovaas, 1987) between 1987 and 1993. Three were enrolled in the University of Oslo Autism Project, a replication site for the UCLA Young Autism Project, between 1989 and 1993. The remaining 2 were enrolled in the University of Kansas Autism Project, which, between 1988 and 1990, was also a replication site for the UCLA Young Autism Project. The section on *Treatment* (see later discussion) presents procedures aimed at ensuring comparability of treatment across sites.

The 10 comparison participants were all enrolled in the UCLA Young Autism Project. Seven had been referred to the project at a time when a study that excluded children with severe mental retardation was in progress (Lovaas, 1987, which ran from 1970 to 1984). Hence, they could not be offered intensive treatment. The remaining 3 participants were referred in 1984 through 1985, when intensive treatment was not being given to any new referrals. None of the comparison participants were treated at the Oslo or Kansas sites because these sites had therapists available to provide intensive treatment to all referrals meeting criteria for inclusion in the present study.

The assignment procedure resulted in two differences in treatment between experimental and control participants: First, experimental participants entered treatment at a later date than did comparison participants (1987 through 1993 vs. 1970 through 1985). McEachin et al. (1993) found that the date at which treatment began did not correlate with outcome for participants in their study. However, some experimental (but not comparison) participants in the present study received treatment *after* the completion of McEachin et al.'s (1993) study. Because of advances in treatment research, the experimental participants may have received a more effective intervention than was available to the comparison participants. This possible discrepancy, rather than the difference in treatment intensity, may account for some or all of the between-group differences in treatment outcome. Second, experimental participants were treated at three different sites, whereas all comparison participants were treated at UCLA. Extensive efforts were made to ensure comparability of personnel across sites, as discussed in the next section. However, no objective data were collected on quality of treatment. Therefore, the differences between experimental and comparison participants with respect to the date of treatment onset and the treatment site are potential confounds.

Treatment

All children in both groups received behavioral treatment, using procedures described in detail in a manual (Lovaas et al., 1981) and associated videotapes (Lovaas & Leaf, 1981). In brief, the treatment was designed to progress gradually and systematically from relatively simple tasks, such as responding to basic requests made by an adult, to more complex skills, such as conversing and making friends with peers. It emphasized the implementation of experimentally validated teaching approaches (cf. Newsom & Rincover, 1989; Schreibman, 1988; Smith, 1993), based on operant conditioning principles, such as shaping, chaining, discrimination training, and contingency management. In the early stages of treatment, instruction took place in a one-to-one discrete trial format, which enabled therapists to devote highly individualized attention to each child. Later, the focus shifted to helping children adjust to group settings, such as classrooms. Children in the experimental group were kept out of school until they reached this stage of treatment. However, all children in the comparison group entered school as soon as they became eligible for special education services (i.e., when they turned 3 years old).

The treatment personnel consisted of student therapists, project directors, clinic supervisors, and senior therapists. At the UCLA and University of Kansas sites, the student therapists were undergraduate and graduate students earning course credit. At the University of Oslo site, they were teacher aides working in the public schools. Student therapists from all sites received extensive training and supervision from the project directors (i.e., the authors, who are psychologists with extensive research and clinical experience in behavioral treatment for children with developmental disabilities). Additional supervision was provided by clinic supervisors (masters' level personnel who had at least 2 years of experience with the treatment model described here) and senior therapists (student therapists with a minimum of 6 months experience and a record of outstanding performance as judged by the clinic supervisors and the authors). Lovaas (1987) described these training and supervision procedures in more detail.

To ensure that the treatment personnel at the Oslo and Kansas sites were comparable to those at the UCLA site, the site directors were required to complete a full-time internship at UCLA prior to data collection. The internship ranged in length from 4 to 9 months of full-time work, depending on the intern's prior experience in one-to-one behavioral instruction. Four other staff members at the Oslo site and 3 others at the Kansas site completed 2- to 3-month full-time internships at UCLA. In addition, during data collection, the site directors were required to meet the following requirements: (a) keep up-to-date on recent findings by attending twice-yearly meetings with other site directors, (b) ensure that each subject received treatment 50 weeks or more per year, (c) attend weekly meetings with each subject and the subject's team (parents, student therapists, and clinic supervisor) in order to provide supervision to the team, and (d) permit site visits from UCLA personnel to evaluate the quality of treatment.

Children in the experimental group received 30 hours per week of instruction in their homes and communities from treatment teams consisting of 4 to 6 student therapists and a senior therapist. Children in the comparison group received 10 hours per week or less of the same instruction (i.e., instruction in their homes and communities that was based on the manual by Lovaas et al., 1981, and was provided by 2 to 4 student therapists and a senior therapist). In both groups, parents were also considered to be important members of their child's treatment team. They were asked to set aside 5 to 10 hours of the time that student therapists or senior therapists were present so that they could work alongside these therapists and become proficient in behavioral teaching

approaches. They were also asked to participate in all team meetings in which decisions were made about their child's instructional program. They were encouraged to use the remaining time that therapists worked with their child to attend to their own needs or those of other family members.

Assessment

All children in the study were tested at intake and follow-up on the Bayley Scales of Infant Development. Examiners, who were blind to the purpose of the study, had extensive experience in administering this test to preschool children with developmental disabilities. At the UCLA and Kansas sites, all of the examiners were either licensed psychologists or advanced doctoral students who were receiving supervision from licensed psychologists. At the Norway site, all children received a minimum of three test administrations at intake and again at follow-up. At least one administration at both intake and follow-up was given by a licensed health care professional who was not working on the project. The highest score obtained at intake and the lowest score obtained at follow-up were used in the data analyses in order to provide a conservative estimate of improvements made by the children who underwent multiple assessments.

Additional data on children's pre- and posttreatment functioning were derived from intake and termination reports, which summarized behavior observations and parent descriptions of each child's functioning. The behavior observations were obtained from logs of children's treatment; these logs adhered to a standard format and were based on data recorded by therapists during treatment sessions. The parent descriptions were obtained from standardized interviews with the site director. (The formats for logs and interviews are available from the first author.)

Reports for the 16 participants treated at the UCLA site were rated by undergraduate students. These students were blind to the purpose of the study and had a minimum of one semester of experience in conducting one-to-one treatment with preschool children with developmental disabilities. Two students independently scored reports for 4 of the 16 children (25%) in order to assess interrater reliability. These reports were selected randomly. A clinic supervisor and project director scored both reports for each of the 5 children seen at the University of Oslo and University of Kansas sites. All raters were trained by (a) having a one-hour meeting with the first author to review definitions of behaviors to be scored and (b) scoring reports for three children who did not participate in the present study. During this training, each rater achieved 90% agreement with another rater and with the first author.

The raters used a scoring system developed by Lovaas (1987). One score was for the highest level of expressive speech displayed by the subject at the time of the evaluation: none, echolalia only, single words, phrases, simple sentences learned by rote, original sentences formulated by the child and used in conversation with others, and apparently typical speech. The definition of *expressive speech* was modified so that not only vocal speech (as in Lovaas, 1987) but also other forms of communication (sign language or picture communication) were assessed.

Additional scores indicated the presence or absence of the following six problems: (a) *apparent sensory deficits* (not responding to spoken language despite normal hearing, not responding to painful stimuli such as ear infections, or not responding to visual stimuli such as a hand being waved within 15.24 cm of the child's eyes); (b) *adult rejection* (arching the back or running away to avoid physical affection most of the time that caregivers offer it); (c) *no toy play*, with *toy play* defined as using an object as intended (e.g., stacking instead of lining blocks, bouncing instead of rubbing a ball, or

pushing a truck instead of spinning the wheels); (d) *stereotyped behaviors,* including (i) repetitive actions lasting 3 or more seconds (such as flicking the fingers in front of the eyes, spinning objects, picking up and dropping objects, rubbing fingers back and forth, running in circles, or repeating a certain sound or word; but not rocking the body or flapping the hands because young, typically developing children commonly display these behaviors), (ii) unusual positions lasting 3 or more seconds (e.g., toe-walking, assuming a posture such as walking with the back hunched over, or gazing out of the corner of the eyes), or (iii) rituals (e.g., insisting that a certain route be followed home, requiring that household furniture be arranged in a certain way, or requiring that daily activities occur in a particular sequence); (e) *tantrums* (an average of one or more instances per day of aggression toward self, others, or property); (f) *no toilet training,* with *toilet training* defined as being out of diapers and having no more than two accidents per week on average.

Results

Initial Analyses

For 9 of the 21 participants (6 of the 11 in the experimental group, 3 of the 10 in the control group), data were obtained on interrater reliability for scoring communicative speech and behavior problems (see preceding section). The raters agreed in 122 of 126 opportunities (9 participants × 2 reports per subject × 7 behaviors per report), or 97%. Thus, interrater reliability appeared satisfactory.

Intake Data

At intake, experimental group children averaged 36 months of age (standard deviation [*SD*] = 6.90) and comparison children averaged 38 months (*SD* = 5.40), a nonsignificant difference. Moreover, their IQs were closely matched, with means differing by only one point: 28 (*SD* = 4.90) and 27 (*SD* = 5.40), respectively (also a nonsignificant difference). No child in either group spoke in words. Table 1 shows the percentage of children in each group who evinced various behavior problems, all of which were common among children in both groups. The groups were about equally likely to display most of these behavior problems. However, fewer experimental than control participants manifested sensory deficits and rejection of adults. These differences were not statistically significant, but the statistical analyses had low power because of the small number of participants.

Table 1
Speech and Behavior Problems (in %) at Intake and Follow-up by Group

	Group	
Time/Problem	Experimental	Comparison
Intake		
Speech (no speech)	100	100
Sensory deficits	45	90
Adult rejection	36	90
No toy play	64	80
Self-stimulation	91	100
Tantrums	91	70
No toilet training	100	100
Follow-up		
Speech (none)	9	100
Echolalia	0	—
Words	73	—
Phrases	9	—
Sentences	9	—
Conversation	9	—
Typical	0	—
Behavior		
Sensory deficit	0	10
Adult rejection	9	30
No toy play	9	30
Self-stimulation	91	100
Tantrums	45	40
No toilet training	45	60

Overall, the absence of statistically significant differences between groups on any intake measure suggests that the subject assignment procedures may have produced similar groups. Nevertheless, one cannot rule out the possibility that important differences existed (especially in the rate of some behavior problems) but escaped detection because of low statistical power. The mean IQ of the groups (28),

the lack of speech of all participants, and the prevalence of behavior problems indicates that, as intended, the criteria for selecting participants yielded a sample that functioned in the range of severe mental retardation and had many autistic behaviors.

Follow-Up Data

All 21 children in the experimental and control groups completed treatment and participated in the follow-up evaluation. The experimental group children were 71 months of age at follow-up (SD = 25.50), and the comparison children, 64 months (SD = 21.22). Although this is not a significant difference, it is noteworthy that the SD of each group was larger than it had been at intake. Follow-up evaluations for most children took place at the end of treatment or shortly thereafter, when the children were 5 to 7 years old. However, evaluations for 4 children (2 in the experimental group and 2 in the control group) took place 3 to 4 years later, when the children were 9 to 10 years old. As a result, the age range of children at follow-up was larger than at intake, when all children were required to be under 46 months of age.

The mean IQ of the experimental group increased from 28 at intake to 36 (SD = 13.14) at follow-up. By contrast, the mean IQ of the comparison group decreased from 27 to 24 (SD = 8.23). At follow-up, the difference in IQ between groups was statistically significant, $t(20) = 2.30$, $p < .05$.

Speech was another variable that revealed a statistically significant difference between groups at follow-up, $U(10, 11) = 6$, $p < .001$. Though no child had spoken in words at intake, 10 of the 11 children in the experimental group did so at the follow-up evaluation compared to 2 of the 10 children in the comparison group. Moreover, the children in the experimental group used their words to label objects and express needs rather than simply echoing what others said (a problem exhibited by many children with autistic behaviors). In addition, 2 of these children progressed past the point of using single words to speaking in phrases and/or sentences. The groups did not differ significantly from each other on any of the behavior problems listed in Table 1.

Individual Data

Table 2 presents data on each subject in the study. Large individual differences are evident among children in the experimental group. Most of these children showed some improvement in IQs, speech, and behavior problems. Nevertheless, some of them did not conform to this pattern. At one extreme, for example, was a child whose IQ remained essentially unchanged, who failed to develop any speech, and whose behavior problems continued unabated in all of the areas listed in Table 2 (Evan). At the other extreme was a child who gained 42 IQ points (going from 17 to 59), developed conversational speech, and exhibited reductions in all behavior problems except self-stimulation (Ken Robert). Thus, children in the experimental group varied substantially in their responsivity to treatment and those in the control group showed less variation. No child made large gains in IQ, and only 2 acquired expressive speech, though most displayed fewer behavior problems. Also of note are the data from the 2 children in each group who were assessed 3 to 4 years after the end of treatment (see Table 2). Their results did not noticeably differ from those of the other participants.

Additional Assessments

Additional test data are available for some children in the experimental group, though not for any children in the comparison group. The following are standard scores at intake and follow-up for composite performance on the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984): Bendik—39, 39, respec-

Table 2
Characteristics of Children at Intake (I) and Follow-Up (F)

Group/Child	Site	Ratio IQ		Expressive speech		Sensory deficits		Adult rejection		No toy play		Stereotype		Tantrums		Not toilet-trained	
		I	F	I	F	I	F	I	F	I	F	I	F	I	F	I	F
Experimental																	
Bendik	Oslo	30	32	none	words	–	–	–	–	–	–	+	+	+	–	+	+
Ken Robert	Oslo	17	59	none	convers.	–	–	–	–	+	–	+	+	+	–	+	–
Petter	Oslo	30	52	none	phases	–	–	–	–	+	–	+	+	+	+	+	–
Nathan	Kansas	30	37	none	words	–	–	–	–	+	–	+	+	+	–	+	–
Dustin[a]	Kansas	31	29	none	words	+	–	–	–	+	–	+	+	+	+	+	–
Luca	UCLA	33	38	none	words	+	–	–	–	–	–	+	+	+	+	+	–
Brian	UCLA	30	43	none	words	+	–	–	–	–	–	–	–	+	–	+	–
Nate H.[a]	UCLA	30	11	none	words	–	–	+	–	+	–	+	+	+	+	+	+
Evan	UCLA	22	31	none	none	–	–	+	+	+	+	+	+	+	+	+	+
Andrew	UCLA	30	16	none	words	+	–	+	–	+	–	+	+	+	–	+	+
Andreas	UCLA	23	46	none	words	+	–	+	–	–	–	+	+	+	–	+	+
Comparison																	
Adrian	UCLA	34	20	none	none	+	–	+	–	+	+	+	+	+	–	+	–
Tameka	UCLA	21	21	none	none	+	+	+	–	+	+	+	+	+	–	+	–
Ryan[a]	UCLA	19	10	none	none	+	–	+	–	+	–	+	+	–	–	+	+
David	UCLA	30	38	none	none	+	–	+	–	–	–	+	+	+	+	+	–
Scott	UCLA	22	15	none	none	+	–	+	–	+	–	+	+	+	–	+	+
Dane	UCLA	30	30	none	none	+	–	+	+	+	–	+	+	+	+	+	+
Edward	UCLA	30	23	none	none	–	–	–	+	+	+	+	+	–	+	+	+
Chris	UCLA	29	31	none	words	+	–	+	+	–	–	+	+	+	–	+	-
Rene[a]	UCLA	34	23	none	none	+	–	+	–	+	–	+	+	+	+	+	+
Shannon	UCLA	24	29	none	words	+	–	+	–	+	–	+	+	–	–	+	+

Note. + = present, – = absent.
[a]Follow-up evaluation took place 3 to 4 years posttreatment.

tively; Ken Robert—41, 80; Petter—63, 63; Dustin—35 (follow-up only); Luca—50, 45; Brian—54, 53; Evan—55, 30; Andreas—55 (follow-up only). On the comprehension section of the Reynell Developmental Language Scales (Reynell & Gruber, 1990), the following age equivalents (in months) were obtained at intake and follow-up: Luca—12, 12; Evan—12, 16; Andreas—12, 24. On the expressive section of this test, all 3 participants obtained an age equivalent of 12 months at both intake and follow-up, though the test protocols indicated that they did not state any words at intake but did do so at follow-up. On the Merrill-Palmer Scale of Mental Tests, Luca, Brian, and Andreas were all untestable at intake, but obtained ratio IQs of 75, 71, and 63, respectively, at follow-up.

The scores on the Vineland are generally consistent with the results presented in Table 2. Ken Robert, who showed significant improvements (see Table 2) also made significant improvements on the Vineland. Evan, who did not show improvement, declined in his score on the Vineland. The other participants, who achieved some improvements in IQ and reductions of behavior problems, obtained scores at follow-up that were close to those obtained at intake. However, no improvement on the Reynell were found in expressive speech (see Table 1) for Luca and Andreas. By contrast, results on the Merrill-Palmer indicated substantial improvements for these children as well as for Evan in visual–spatial skills (a variable not shown in Table 2).

Clinical Significance

For commonly used measures such as IQ and language development, investigators may determine clinical significance either by simply comparing group averages or by conducting statistical analyses of effect size (cf. Cohen, 1990). With regard to IQ, the 12-point advantage of the experimental group over the comparison group is comparable to results achieved in early intervention programs for substantially less impaired populations than the one in this study. Advantages of 10±2 IQ points are typically found in such studies and are generally regarded as having practical importance (e.g., Weinberg, 1989). Consistent with the interpretation that the difference between groups is substantial, correlation of group assignment with follow-up IQs as a measure of effect size (Rosnow & Rosenthal, 1988) yielded an r of .48, which is a moderately large effect for a study on a psychoeducational treatment. The finding that 10 of 11 experimental participants used words to communicate, compared to only 2 of 10 comparison participants, is also indicative of a substantial effect.

However, the 12-point IQ advantage of the experimental group over the control group is substantially smaller than the 31-point advantage reported for children with autism and with IQs above 35 who received the treatment given to the experimental group (Lovaas, 1987). Also, the treatment did not enable any of the participants to perform in the average range of intelligence (IQ > 70). (This index of clinical significance was advocated by Jacobson and Truax, 1991, though they noted that the index may be unrealistically stringent for severely delayed populations.) Overall, then, these analyses indicate that treatment was associated with substantial improvements. However, the improvements were much more modest than those reported for higher functioning children who received the same treatment, and children remained very much delayed in development.

Discussion

In this study we analyzed archival data on outcomes attained by 11 preschool children with both severe mental retardation and autistic features who received intensive behavioral treatment (30 hours per week of one-to-one, in-home instruction for 2 years). These intensively treated children were compared to 10 similar

children who received minimal treatment. At the follow-up evaluation, the intensively treated children achieved a higher mean IQ (36 vs. 24) and demonstrated more expressive speech than did the minimally treated children. Ten of the 11 intensively treated children spoke in words and used them to label objects and express needs at follow-up, whereas only 2 of 10 minimally treated children did so. Although these gains are much less than those made by higher functioning children who have received the same treatment (Lovaas, 1987), they may be greater than what investigators would have expected from this often-neglected population (cf. Guralnick & Bricker, 1987) and appear to be clinically meaningful, as discussed in the preceding section.

Intensively treated children did not differ significantly from minimally treated children in the prevalence of behavior problems at follow-up. Several interpretations of this finding are possible: (a) Though minimal treatment was not enough to raise children's IQs and establish communicative speech, it may have sufficed to alleviate behavior problems. (b) Behavior problems may have been reduced by the special education classes in which the minimally treated children (but not the intensively treated children) were enrolled. (c) Behavior problems may have diminished because of maturational changes that occurred independently of the services the children received. Or (d) as a result of the small number of children in the study, the statistical analyses may have had too little power to detect differences between intensively treated and minimally treated children.

The present study contained several methodological features that may enhance the credibility of the findings. First, we evaluated an intervention that combined treatment approaches that have been experimentally validated by numerous investigators. Hence, the intervention has been developed systematically and warrants evaluation in a long-term outcome study such as this. Second, the intervention has been detailed in a manual (Lovaas et al., 1981) and videotapes (Lovaas & Leaf, 1981), which are likely to facilitate replications. Third, it contained a comparison group of minimally treated children who appeared to be similar to the intensively treated children prior to treatment. Consequently, the gains made by the intensively treated children may be attributable to the intervention they received rather than some other factor. Finally, the assessment instruments centered on variables that have clear clinical relevance and were administered by examiners who were not aware of the purpose of the study.

On the other hand, because this study was a retrospective analysis rather than a planned investigation, it contains a number of limitations. Perhaps the most serious of these is that children were assigned to the intensive treatment group or the minimal treatment group based on therapist availability instead of a more arbitrary procedure, such as alternating referrals (assigning first referral to the intensive treatment group, the next to the minimal treatment group, the third to the intensive treatment group, and so on). Although no evidence indicates that this procedure biased how groups were formed and although such a procedure is common and generally accepted in clinical research (Kazdin, 1992), a more arbitrary assignment procedure would have been sounder. This is particularly true in the present study because the assignment procedure resulted in a discrepancy between when experimental and comparison participants were treated. (Experimental participants were treated from 1987 to 1993; control participants, from the early 1970s to 1987; see *Design* section). This discrepancy may have resulted in differences in treatment quality across groups. Further concerns may be raised about (a) the small sample size; (b) the reliability of the data on behavior problems (which were based on parent report and treatment logs rather than direct observation by the raters); (c) the statistically

insignificant but perhaps clinically important difference between groups in the prevalence of behavior problems at intake (children in the comparison group had more frequent problems on some measures); (d) the omission of variables that might have given a more complete picture of the results, such as demographic data on the children's families, information on other treatments the children may have received, adaptive behavior inventories, standardized measures of speech and behavior problems, and evaluations of how treatment influenced overall family functioning; and (e) the absence of direct measures of the quality and quantity of treatment provided to children. In view of these rather substantial weaknesses, the results need to be interpreted cautiously despite the methodological safeguards just identified.

A difficult issue raised by results of the present study is whether it is appropriate to expend such a large amount of time and effort (3,000 hours or more of one-to-one instruction per child) in order to achieve the gains typically made by the intensively treated children. Most intensively treated children showed improvements in IQ and speech but remained quite delayed in these and other areas. As previously discussed, many other important areas of functioning were not assessed, and it is possible that evaluations in these areas would have yielded a more favorable result. Also of interest is that substantial individual differences emerged between children. Because of these differences, it would be useful to develop ways to identify the most treatment-responsive children and to focus treatment on these children. As for the treatment itself, a number of improvements seem possible. For example, the intensively treated children received 30 hours per week of treatment rather than 40, as has been recommended elsewhere (e.g., Lovaas, 1987). Perhaps the provision of 40 hours per week would have yielded larger gains. Also, the manual on which the intensive treatment was based (Lovaas et al., 1981) has become dated in some respects because of the large body of research that has appeared since its publication. A revision of the manual is underway. Perhaps basing treatment on the revised manual would enhance children's outcomes. In addition, other models of behavioral intervention, several of which have yielded encouraging results in preliminary investigations (Green, 1996), might be more effective than the model used in the present study. Thus, further research on intensive treatment for preschoolers with both severe mental retardation and autistic features seems warranted.

References

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

Anderson, S. R., Avery, D. L., DiPietro, E. K., Edwards, G. L., & Christian, W. P. (1987). Intensive home-based intervention with autistic children. *Education and Treatment of Children, 10,* 352–366.

Bayley, N. (1969). *Bayley Scales of Infant Development.* New York: Psychological Corp.

Birnbrauer, J. S., & Leach, D. J. (1993). The Murdoch Early Intervention Program after two years. *Behaviour Change, 10,* 63–74.

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

Fenske, E. C., Zalenski, S., Krantz, P. J., & McClanahan, L. E. (1985). Age at intervention and treatment outcome for autistic children in a comprehensive intervention program. *Analysis and Intervention in Developmental Disabilities, 5,* 49–58.

Green, G. (1996). Early behavioral intervention for autism: What does research tell us? In C. Maurice (Ed.), *Behavioral intervention for young children with autism* (pp. 29–44). Austin, TX: Pro-Ed.

Guralnick, M. J., & Bricker, D. (1987). The effectiveness of early intervention for children with cognitive and general developmental delays. In M. J. Guralnick & F. C. Bennett (Eds.), *The effectiveness of early intervention for at-risk and handicapped children* (pp. 79–108). New York: Academic Press.

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.

Hoyson, M., Jamieson, B., & Strain, P. S. (1984). Individualized group instruction of normally developing and autistic-like children: A description and evaluation of the LEAP curriculum model. *Journal of the Division of Early Childhood, 8,* 157–181.

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

Kazdin, A. E. (1992). *Research design in clinical psychology* (2nd ed.). Needham Heights, MA: Allyn & Bacon.

Lord, C., & Schopler, E. (1988). Intellectual and developmental assessment of autistic children from preschool to schoolage: Clinical implications of two follow-up studies. In E. Schopler & G. B. Mesibov (Eds.), *Diagnosis and assessment in autism* (pp. 167–181). New York: Plenum.

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., & Leaf, R. B. (1981). *Five video tapes for teaching developmentally disabled children.* Baltimore: University Park Press.

Lovaas, O. I., & Smith, T. (1988). Intensive behavioral treatment for young autistic children. In B. B. Lahey & A. E. Kazdin (Eds.), *Advances in clinical child psychology* (Vol. 11, pp. 285–324). New York: Plenum.

McEachin, J. J., Smith, T., & Lovaas, O. I. (1993). Long-term outcome for children with autism who received early intensive behavioral treatment. *American Journal on Mental Retardation, 97,* 359–372.

Newsom, C., & Rincover, A. (1989). Autism. In E. J. Mash & R. A. Barklay (Eds.), *Treatment of childhood disorders* (pp. 286–346). New York: Guilford.

Reynell, J. K., & Gruber, C. P. (1990). *Reynell Developmental Language Scales.* Los Angeles: Western Psychological Services.

Rosnow, R. L., & Rosenthal, R. (1988). Focused tests of significance and effect size estimation in counseling psychology. *Journal of Counseling Psychology, 35,* 203–208.

Schreibman, L. (1988). *Autism.* Beverly Hills, CA: Sage.

Simeonnson, R. J., Olley, J. G., & Rosenthal, S. L. (1987). Early intervention for children with autism. In M. J. Guralnick & F. C. Bennett (Eds.), *The effectiveness of early intervention for at-risk and handicapped children* (pp. 275–296). Orlando, FL: Academic Press.

Smith, T. (1993). Autism. In T. R. Giles (Ed.), *Effective psychotherapies* (pp. 107–133). New York: Plenum.

Smith, T., McEachin, J. J., & Lovaas, O. I. (1993). Comments on replication and evaluation of outcome. *American Journal on Mental Retardation, 97,* 385–391.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales: Interview edition: Survey form manual.* Circle Pines, MN: American Guidance Service.

Weinberg, R. A. (1989). Intelligence and IQ: Landmark issues and great debates. *American Psychologist, 44,* 98–104.

Wing, L. (1981). Language, social, and cognitive impairments in autism and severe mental retardation. *Journal of Autism and Developmental Disorders, 9,* 11–29.

Received 4/14/95, first decision 10/16/95, second decision 8/30/96, third decision 12/10/96, accepted 1/9/97.

The authors thank Kristen Villone for her assistance in data collection. Requests for reprints should be sent to Tristram Smith, Department of Psychology, Washington State University, Pullman, WA 99164-4820.