Journal of Consulting and Clinical Psychology
1987, Vol. 55, No. 1, 3-9

0022-006X/87/$00.75

Behavioral Treatment and Normal Educational and Intellectual Functioning in Young Autistic Children

O. Ivar Lovaas
University of California, Los Angeles

Autism is a serious psychological disorder with onset in early childhood. Autistic children show minimal emotional attachment, absent or abnormal speech, retarded IQ, ritualistic behaviors, aggression, and self-injury. The prognosis is very poor, and medical therapies have not proven effective. This article reports the results of behavior modification treatment for two groups of similarly constituted, young autistic children. Follow-up data from an intensive, long-term experimental treatment group (n = 19) showed that 47% achieved normal intellectual and educational functioning, with normal-range IQ scores and successful first grade performance in public schools. Another 40% were mildly retarded and assigned to special classes for the language delayed, and only 10% were profoundly retarded and assigned to classes for the autistic/retarded. In contrast, only 2% of the control-group children (n = 40) achieved normal educational and intellectual functioning; 45% were mildly retarded and placed in language-delayed classes, and 53% were severely retarded and placed in autistic/retarded classes.

Kanner (1943) defined autistic children as children who exhibit (a) serious failure to develop relationships with other people before 30 months of age, (b) problems in development of normal language, (c) ritualistic and obsessional behaviors ("insistence on sameness"), and (d) potential for normal intelligence. A more complete behavioral definition has been provided elsewhere (Lovaas, Koegel, Simmons, & Long, 1973). The etiology of autism is not known, and the outcome is very poor. In a follow-up study on young autistic children, Rutter (1970) reported that only 1.5% of his group (n = 63) had achieved normal functioning. About 35% showed fair or good adjustment, usually required some degree of supervision, experienced some difficulties with people, had no personal friends, and showed minor oddities of behavior. The majority (more than 60%) remained severely handicapped and were living in hospitals for mentally retarded or psychotic individuals or in other protective settings. Initial IQ scores appeared stable over time. Other studies (Brown, 1969; DeMyer et al., 1973; Eisenberg, 1956; Freeman, Ritvo, Needleman, & Yokota, 1985; Havelkova, 1968) report similar data. Higher scores on IQ tests, communicative speech, and appropriate play are considered to be prognostic of better outcome (Lotter, 1967).

Medically and psychodynamically oriented therapies have not proven effective in altering outcome (DeMyer, Hingtgen, & Jackson, 1981). No abnormal environmental etiology has been identified within the children's families (Lotter, 1967). At present, the most promising treatment for autistic persons is behavior modification as derived from modern learning theory (DeMyer et al., 1981). Empirical results from behavioral intervention with autistic children have been both positive and negative. On the positive side, behavioral treatment can build complex behaviors, such as language, and can help to suppress pathological behaviors, such as aggression and self-stimulatory behavior. Clients vary widely in the amount of gains obtained but show treatment gains in proportion to the time devoted to treatment. On the negative side, treatment gains have been specific to the particular environment in which the client was treated, substantial relapse has been observed at follow-up, and no client has been reported as recovered (Lovaas et al., 1973).

The present article reports a behavioral-intervention project (begun in 1970) that sought to maximize behavioral treatment gains by treating autistic children during most of their waking hours for many years. Treatment included all significant persons in all significant environments. Furthermore, the project focused on very young autistic children (below the age of 4 years) because it was assumed that younger children would be less likely to discriminate between environments and therefore more likely to generalize and to maintain their treatment gains. Finally, it was assumed that it would be easier to successfully mainstream a very young autistic child into preschool than it would be to mainstream an older autistic child into primary school.

It may be helpful to hypothesize an outcome of the present study from a developmental or learning point of view. One may assume that normal children learn from their everyday environ-

This study was supported by Grant MH-11440 from the National Institute of Mental Health. Aspects of this study were presented at the 1982 convention of the American Psychological Association, Washington, DC, by Andrea Ackerman, Paula Firestone, Gayle Goldstein, Ronald Leaf, John McEachin, and the author. The author expresses his deep appreciation to the many undergraduate students at the University of California, Los Angeles, who served as student therapists on the project, to the many graduate students who served as clinic supervisors, and to the many parents who trusted their children to our care. Special thanks to Laura Schreibman and Robert Koegel, who collaborated in the early stages of this research project. Donald Baer, Bruce Baker, Bradley Bucher, Arthur Woodward, and Haikang Shen provided statistical advice and help in manuscript preparation. B. J. Freeman's help in arranging access to Control Group 2 data is also appreciated.

Correspondence concerning this article should be addressed to O. Ivar Lovaas, Psychology Department, University of California, 405 Hilgard Avenue, Los Angeles, California 90024.

ments most of their waking hours. Autistic children, conversely, do not learn from similar environments. We hypothesized that construction of a special, intense, and comprehensive learning environment for very young autistic children would allow some of them to catch up with their normal peers by first grade.

Method

Subjects

Subjects were enrolled for treatment if they met three criteria: (a) independent diagnosis of autism from a medical doctor or a licensed PhD psychologist, (b) chronological age (CA) less than 40 months if mute and less than 46 months if echolalic, and (c) prorated mental age (PMA) of 11 months or more at a CA of 30 months. The last criterion excluded 15% of the referrals.

The clinical diagnosis of autism emphasized emotional detachment, extreme interpersonal isolation, little if any toy or peer play, language disturbance (mutism or echolalia), excessive rituals, and onset in infancy. The diagnosis was based on a structured psychiatric interview with parents, on observations of the child's free-play behaviors, on psychological testing of intelligence, and on access to pediatric examinations. Over the 15 years of the project, the exact wording of the diagnosis changed slightly in compliance with changes in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM–III; American Psychiatric Association, 1980). During the last years, the diagnosis was made in compliance with DSM–III criteria (p. 87). In almost all cases, the diagnosis of autism had been made prior to family contact with the project. Except for one case each in the experimental group and Control Group 1, all cases were diagnosed by staff of the Department of Child Psychiatry, University of California, Los Angeles (UCLA) School of Medicine. Members of that staff have contributed to the writing of the DSM–III and to the diagnosis of autism adopted by the National Society for Children and Adults with Autism. If the diagnosis of autism was not made, the case was referred elsewhere. In other words, the project did not select its cases. More than 90% of the subjects received two or more independent diagnoses, and agreement on the diagnosis of autism was 100%. Similarly high agreement was not reached for subjects who scored within the profoundly retarded range on intellectual functioning (PMA < 11 months); these subjects were excluded from the study.

Treatment Conditions

Subjects were assigned to one of two groups: an intensive-treatment experimental group (n = 19) that received more than 40 hours of one-to-one treatment per week, or the minimal-treatment Control Group 1 (n = 19) that received 10 hours or less of one-to-one treatment per week. Control Group 1 was used to gain further information about the rate of spontaneous improvement in very young autistic children, especially those selected by the same agency that provided the diagnostic work-up for the intensive-treatment experimental group. Both treatment groups received treatment for 2 or more years. Strict random assignment (e.g., based on a coin flip) to these groups could not be used due to parent protest and ethical considerations. Instead, subjects were assigned to the experimental group unless there was an insufficient number of staff members available to render treatment (an assessment made prior to contact with the family). Two subjects were assigned to Control Group 1 because they lived further away from UCLA than a 1-hr drive, which made sufficient staffing unavailable to those clients. Because fluctuations in staff availability were not associated in any way with client characteristics, it was assumed that this assignment would produce unbiased groups. A large number of pretreatment measures were collected to test this assumption. Subjects did not change group assignment. Except for two families who left the experimental group within the first 6 months (this group began with 21 subjects), all families stayed with their groups from beginning to end.

Assessments

Pretreatment mental age (MA) scores were based on the following scales (in order of the frequency of their use): the Bayley Scales of Infant Development (Bayley, 1955), the Cattell Infant Intelligence Scale (Cattell, 1960), the Stanford-Binet Intelligence Scale (Thorndike, 1972), and the Gesell Infant Development Scale (Gesell, 1949). The first three scales were administered to 90% of the subjects, and relative usage of these scales was similar in each group. Testing was carried out by graduate students in psychology who worked under the supervision of clinical psychologists at UCLA or licensed PhD psychologists at other agencies. The examiner chose the test that would best accommodate each subject's developmental level, and this decision was reached independently of the project staff. Five subjects were judged to be untestable (3 in the experimental group and 2 in Control Group 1). Instead, the Vineland Social Maturity Scale (Doll, 1953) was used to estimate their MAs (with the mother as informant). To adjust for variations in MA scores as a function of the subject's CA at the time of test administration, PMA scores were calculated for a CA at 30 months (MA/CA × 30).

Behavioral observations were based on videotaped recordings of the subject's free-play behavior in a playroom equipped with several simple early-childhood toys. These videotaped recordings were subsequently scored for amount of (a) *self-stimulatory behaviors,* defined as prolonged ritualistic, repetitive, and stereotyped behavior such as body-rocking, prolonged gazing at lights, excessive hand-flapping, twirling the body as a top, spinning or lining of objects, and licking or smelling of objects or wall surfaces; (b) *appropriate play behaviors,* defined as those limiting the use of toys in the playroom to their intended purposes, such as pushing the truck on the floor, pushing buttons on the toy cash register, putting a record on the record player, and banging with the toy hammer; and (c) *recognizable words,* defined to include any recognizable word, independent of whether the subject used it in a meaningful context or for communicative purposes. One observer who was naive about subjects' group placement scored all tapes after being trained to agree with two experienced observers (using different training tapes from similar subjects). Interobserver reliability was scored on 20% of the tapes (randomly selected) and was computed for each category of behavior for each subject by dividing the sum of observer agreements by the sum of agreements and disagreements. These scores were then summed and averaged across subjects. The mean agreement (based both on occurrences and nonoccurrences) was 91% for self-stimulatory behavior, 85% for appropriate play behavior, and 100% for recognizable words. A more detailed description of these behavioral recordings has been provided elsewhere (Lovaas et al., 1973).

A 1-hr parent interview about the subjects' earlier history provided some diagnostic and descriptive information. Subjects received a score of 1 for each of the following variables parents reported: no recognizable words; no toy play (failed to use toys for their intended function); lack of emotional attachment (failed to respond to parents' affection); apparent sensory deficit (parents had suspected their child to be blind or deaf because the child exhibited no or minimal eye contact and showed an unusually high pain threshold); no peer play (subject did not show interactive play with peers); self-stimulatory behavior; tantrums (aggression toward family members or self); and no toilet training. These 8 measures from parents' intake interviews were summed to provide a sum pathology score. The intake interview also provided information about abnormal speech (0 = normal and meaningful language, however limited; 1 = echolalic language used meaningfully [e.g., to express needs]; 2 = echolalia; and 3 = mute); age of walking; number of siblings in the family; socioeconomic status of the father; sex; and neurological examinations (including EEGs and CAT scans) that resulted in findings of pathology. Finally, CA at first diagnosis and at the beginning of the

present treatment were recorded. This yielded a total of 20 pretreatment measures, 8 of which were collapsed into 1 measure (sum pathology).

A brief clinical description of the experimental group at intake follows (identical to that for Control Group 1): Only 2 of the 19 subjects obtained scores within the normal range of intellectual functioning; 7 scored in the moderately retarded range, and 10 scored in the severely retarded range. No subject evidenced pretend or imaginary play, only 2 evidenced *complex* (several different or heterogeneous behaviors that together formed one activity) play, and the remaining subjects showed *simple* (the same elementary but appropriate response made repeatedly) play. One subject showed minimal appropriate speech, 7 were echolalic, and 11 were mute. According to the literature that describes the developmental delays of autistic children in general, the autistic subjects in the present study constituted an average (or below average) sample of such children.

Posttreatment measures were recorded as follows: Between the ages of 6 and 7 years (when a subject would ordinarily have completed first grade), information about the subjects' first-grade placement was sought and validated; about the same time, an IQ score was obtained. Testing was carried out by examiners who were naive about the subjects' group placement. Different scales were administered to accommodate different developmental levels. For example, a subject with a regular educational placement received a Wechsler Intelligence Scale for Children–Revised (WISC–R; Wechsler, 1974) or a Stanford-Binet Intelligence Scale (Thorndike, 1972), whereas a subject in an autistic/retarded class received a nonverbal test like the Merrill-Palmer Pre-School Performance Test (Stutsman, 1948). In all instances of subjects having achieved a normal IQ score, the testing was eventually replicated by other examiners. The scales (in order of the frequency of usage) included the WISC–R (Wechsler, 1974), the Stanford-Binet (Thorndike, 1972), the Peabody Picture Vocabulary Test (Dunn, 1981), the Wechsler Pre-School Scale (Wechsler, 1967), the Bayley Scales of Infant Development (Bayley, 1955), the Cattell Infant Intelligence Scale (Cattell, 1960), and the Leiter International Performance Scale (Leiter, 1959). Subjects received a score of 3 for *normal functioning* if they received a score on the WISC–R or Stanford-Binet in the normal range, completed first grade in a normal class in a school for normal children, and were advanced to the second grade by the teacher. Subjects received a score of 2 if they were placed in first-grade in a smaller *aphasia* (language delayed, language handicapped, or learning disabled) class. Placement in the aphasia class implied a higher level of functioning than placement in classes for the autistic/retarded, but the diagnosis of autism was almost always retained. A score of 1 was given if the first-grade placement was in a class for the autistic/retarded and if the child's IQ score fell within the severely retarded range.

Treatment Procedure

Each subject in the experimental group was assigned several well trained student therapists who worked (part-time) with the subject in the subject's home, school, and community for an average of 40 hr per week for 2 or more years. The parents worked as part of the treatment team throughout the intervention; they were extensively trained in the treatment procedures so that treatment could take place for almost all of the subjects' waking hours, 365 days a year. A detailed presentation of the treatment procedure has been presented in a teaching manual (Lovaas et al., 1980). The conceptual basis of the treatment was reinforcement (operant) theory; treatment relied heavily on discrimination-learning data and methods. Various behavioral deficiencies were targeted, and separate programs were designed to accelerate development for each behavior. High rates of aggressive and self-stimulatory behaviors were reduced by being ignored; by the use of time-out; by the shaping of alternate, more socially acceptable forms of behavior; and (as a last resort) by the delivery of a loud "no" or a slap on the thigh contingent upon the presence of the undesirable behavior. Contingent physical aversives were not used in the control group because inadequate staffing in that group did not allow for adequate teaching of alternate, socially appropriate behaviors.

During the first year, treatment goals consisted of reducing self-stimulatory and aggressive behaviors, building compliance to elementary verbal requests, teaching imitation, establishing the beginnings of appropriate toy play, and promoting the extension of the treatment into the family. The second year of treatment emphasized teaching expressive and early abstract language and interactive play with peers. Treatment was also extended into the community to teach children to function within a preschool group. The third year emphasized the teaching of appropriate and varied expression of emotions; preacademic tasks like reading, writing, and arithmetic; and *observational learning* (learning by observing other children learn). Subjects were enrolled only in those preschools where the teacher helped to carry out the treatment program. Considerable effort was exercised to mainstream subjects in a normal (average and public) preschool placement and to avoid initial placement in special education classes with the detrimental effects of exposure to other autistic children. This occasionally entailed withholding the subject's diagnosis of autism. If the child became known as autistic (or as "a very difficult child") during the first year in preschool, the child was encouraged to enroll in another, unfamiliar school (to start fresh). After preschool, placement in public education classes was determined by school personnel. All children who successfully completed normal kindergarten successfully completed first grade and subsequent normal grades. Children who were observed to be experiencing educational and psychological problems received their school placement through Individualized Educational Plan (IEP) staffings (attended by educators and psychologists) in accordance with the Education For All Handicapped Children Act of 1975.

All subjects who went on to a normal first grade were reduced in treatment from the 40 hr per week characteristic of the first 2 years to 10 hr or less per week during kindergarten. After a subject had started first grade, the project maintained a minimal (at most) consultant relationship with some families. In two cases, this consultation and the subsequent correction of problem behaviors were judged to be essential in maintaining treatment gains. Subjects who did not recover in the experimental group received 40 hr or more per week of one-to-one treatment for more than 6 years (more than 14,000 hr of one-to-one treatment), with some improvement shown each year but with only 1 subject recovering.

Subjects in Control Group 1 received the same kind of treatment as those in the experimental group but with less intensity (less than 10 hr of one-to-one treatment per week) and without systematic physical aversives. In addition, these subjects received a variety of treatments from other sources in the community such as those provided by small special education classes.

Control Group 2 consisted of 21 subjects selected from a larger group (N = 62) of young autistic children studied by Freeman et al. (1985). These subjects came from the same agency that diagnosed 95% of our other subjects. Data from Control Group 2 helped to guard against the possibility that subjects who had been referred to us for treatment constituted a subgroup with particularly favorable or unfavorable outcomes. To provide a group of subjects similar to those in the experimental group and Control Group 1, subjects for Control Group 2 were selected if they were 42 months old or younger when first tested, had IQ scores above 40 at intake, and had follow-up testing at 6 years of age. These criteria resulted in the selection of 21 subjects. Subjects in Control Group 2 were treated like Control Group 1 subjects but were not treated by the Young Autism Project described here.

Results

Pretreatment Comparisons

Eight pretreatment variables from the experimental group and Control Group 1 (CA at first diagnosis, CA at onset of treat-

Table 1
Means and F Ratios From Comparisons Between Groups on Intake Variables

Group	Diagnosis CA	Treatment CA	PMA	Recognizable words	Toy play	Self-stimulation	Sum pathology	Abnormal speech
Experimental	32.0	34.6	18.8	.42	28.2	12.1	6.9	2.4
Control 1	35.3	40.9	17.1	.58	20.2	19.6	6.4	2.2
F[a]	1.58	4.02*	1.49	.92	2.76	3.37	.82	.36

Note. CA = chronological age; PMA = prorated mental age. Experimental group, $n = 19$; Control Group 1, $n = 19$.
[a] $df = 1, 36$.
* $p < .05$.

ment, PMA, sum pathology, abnormal speech, self-stimulatory behavior, appropriate toy play, and recognizable words) were subjected to a multivariate analysis of variance (MANOVA; Brecht & Woodward, 1984). The means and F ratios from this analysis are presented in Table 1. As can be seen, there were no significant differences between the groups except for CA at onset of our treatment ($p < .05$). Control subjects were 6 months older on the average than experimental subjects (mean CAs of 35 months vs. 41 months, respectively). These differences probably reflect the delay of control subjects in their initiation into the treatment project because of staff shortages; analysis will show that differential CAs are not significantly related to outcome. To ascertain whether another test would reveal a statistically significant difference between the groups on toy play, descriptions of the subjects' toy play (taken from the videotaped recordings) were typed on cards and rated for their developmental level by psychology students who were naive about the purpose of the ratings and subject group assignment. The ratings were reliable among students ($r = .79$, $p < .001$), and an F test showed no significant difference in developmental levels of toy play between the two groups.

The respective means from the experimental group and Control Group 1 on the eight variables from the parent interview were .89 and .74 for sensory deficit, .63 and .42 for adult rejection, .58 and .47 for no recognizable words, .53 and .63 for no toy play, 1.0 and 1.0 for no peer play, .95 and .89 for body self-stimulation, .89 and .79 for tantrums, and .68 and .63 for no toilet training. The experimental group and Control Group 1 were also similar in onset of walking (6 vs. 8 early walkers; 1 vs. 2 late walkers), number of siblings in the family (1.26 in each group), socioeconomic status of the father (Level 49 vs. Level 54 according to 1950 Bureau of the Census standards), boys to girls (16:3 vs. 11:8); and number of subjects referred for neurological examinations (10 vs. 15) who showed signs of damage (0 vs. 1). The numbers of favorable versus unfavorable prognostic signs (directions of differences) on the pretreatment variables divide themselves equally between the groups. In short, the two groups appear to have been comparable at intake.

Follow-Up Data

Subjects' PMA at intake, follow-up educational placement, and IQ scores were subjected to a MANOVA that contrasted the experimental group with Control Groups 1 and 2. At intake, there were no significant differences between the experimental group and the control groups. At follow-up, the experimental group was significantly higher than the control groups on educational placement ($p < .001$) and IQ ($p < .01$). The two control groups did not differ significantly at intake or at follow-up. In short, data from Control Group 2 replicate those from Control Group 1 and further validate the effectiveness of our experimental treatment program. Data are given in Table 2 that show the group means from pretreatment PMA and posttreatment educational placement and IQ scores. The table also shows the F ratios and significance levels of the three group comparisons.

In descriptive terms, the 19-subject experimental group shows 9 children (47%) who successfully passed through normal first grade in a public school and obtained an average or above average score on IQ tests ($M = 107$, range = 94–120). Eight subjects (42%) passed first grade in aphasia classes and obtained a mean IQ score within the mildly retarded range of intellectual functioning ($M = 70$, range = 56–95). Only two children (10%) were placed in classes for autistic/retarded children and scored in the profoundly retarded range (IQ < 30).

There were substantial increases in the subjects' levels of intellectual functioning after treatment. The experimental group subjects gained on the average of 30 IQ points over Control Group 1 subjects. Thus the number of subjects who scored within the normal range of intellectual functioning increased from 2 to 12, whereas the number of subjects within the moderate-to-severe range of intellectual retardation dropped from 10 to 3. As of 1986, the achievements of experimental group sub-

Table 2
Means and F Ratios for Measures at Pretreatment and Posttreatment

Group	Intake PMA	Follow-up EDP	Follow-up IQ
	Means		
Experimental	18.8	2.37	83.3
Control 1	17.1	1.42	52.2
Control 2	17.6	1.57	57.5
	F ratios[a]		
Experimental × Control 1	1.47	23.6**	14.4**
Experimental × Control 2	0.77	17.6**	10.4*
Control 1 × Control 2	0.14	0.63	0.45

Note. PMA = prorated mental age; EDP = educational placement. Experimental group, $n = 19$; Control Group 1, $n = 19$; Control Group 2, $n = 21$.
[a] $df = 1, 56$.
* $p < .01$. ** $p < .001$.

Table 3
Educational Placement and Mean and Range of IQ at Follow-Up

Group	Recovered	Aphasic	Autistic/Retarded
Experimental			
N	9	8	2
M IQ	107	70	30
Range	94–120	56–95	—[a]
Control Group 1			
N	0	8	11
M IQ	—	74	36
Range	—	30–102	20–73
Control Group 2			
N	1	10	10
M IQ	99	67	44
Range	—	49–81	35–54

Note. Dashes indicate no score or no entry.
[a] Both children received the same score.

jects have remained stable. Only 2 subjects have been reclassified: 1 subject (now 18 years old) was moved from an aphasia to a normal classroom after the sixth grade; 1 subject (now 13 years old) was moved from an aphasia to an autistic/retarded class placement.

The MA and IQ scores of the two control groups remained virtually unchanged between intake and follow-up, consistent with findings from other studies (Freeman et al., 1985; Rutter, 1970). The stability of the IQ scores of the young autistic children, as reported in the Freeman et al. study, is particularly relevant for the present study because it reduces the possibility of spontaneous recovery effects. In descriptive terms, the combined follow-up data from the control groups show that their subjects fared poorly: Only 1 subject (2%) achieved normal functioning as evidenced by normal first-grade placement and an IQ of 99 on the WISC–R; 18 subjects (45%) were in aphasia classes (mean IQ = 70, range = 30–101); and 21 subjects (53%) were in classes for the autistic/retarded (mean IQ = 40, range = 20–73). Table 3 provides a convenient descriptive summary of the main follow-up data from the three groups.

One final control procedure subjected 4 subjects in the experimental group (Ackerman, 1980) and 4 subjects in Control Group 1 (McEachin & Leaf, 1984) to a treatment intervention in which one component of treatment (the loud "no" and occasional slap on the thigh contingent on self-stimulatory, aggressive, and noncompliant behavior) was at first withheld and then introduced experimentally. A within-subjects replication design was used across subjects, situations, and behaviors, with baseline observations varying from 3 weeks to 2 years after treatment had started (using contingent positive reinforcement only). During baseline, when the contingent-aversive component was absent, small and unstable reductions were observed in the large amount of inappropriate behaviors, and similar small and unstable increases were observed in appropriate behaviors such as play and language. These changes were insufficient to allow for the subjects' successful mainstreaming. Introduction of contingent aversives resulted in a sudden and stable reduction in the inappropriate behaviors and a sudden and stable increase in appropriate behaviors. This experimental intervention helps to establish two points: First, at least one component in the treatment program functioned to produce change, which helps to reduce the effect of placebo variables. Second, this treatment component affected both the experimental and control groups in a similar manner, supporting the assumption that the two groups contained similar subjects.

Analyses of variance were carried out on the eight pretreatment variables to determine which variables, if any, were significantly related to outcome (gauged by educational placement and IQ) in the experimental group and Control Group 1. Prorated mental age was significantly ($p < .03$) related to outcome in both groups, a finding that is consistent with reports from other investigators (DeMyer et al., 1981). In addition, abnormal speech was significantly ($p < .01$) related to outcome in Control Group 1. Chronological age at onset of our treatment was not related to outcome, which is important because the two groups differed significantly on this variable at intake (by 6 months). The failure of CA to relate to outcome may be based on the very young age of all subjects at onset of treatment.

Conceivably, a linear combination of pretreatment variables could have predicted outcome in the experimental group. Using a discriminant analysis (Ray, 1982) with the eight variables used in the first multivariate analysis, it was possible to predict perfectly the 9 subjects who did achieve normal functioning, and no subject was predicted to achieve this outcome who did not. In this analysis, PMA was the only variable that was significantly related to outcome. Finally, when this prediction equation was applied to Control Group 1 subjects, 8 were predicted to achieve normal functioning with intensive treatment; this further verifies the similarity between the experimental group and Control Group 1 prior to treatment.

Discussion

This article reports the results of intensive behavioral treatment for young autistic children. Pretreatment measures revealed no significant differences between the intensively treated experimental group and the minimally treated control groups. At follow-up, experimental group subjects did significantly better than control group subjects. For example, 47% of the experimental group achieved normal intellectual and educational functioning in contrast to only 2% of the control group subjects.

The study incorporated certain methodological features designed to increase confidence in the effectiveness of the experimental group treatment:

1. Pretreatment differences between the experimental and control groups were minimized in four ways. First, the assignment of subjects to groups was as random as was ethically possible. The assignment apparently produced unbiased groups as evidenced by similar scores on the 20 pretreatment measures and by the prediction that an equal number of Control Group 1 and experimental group subjects would have achieved normal functioning had the former subjects received intensive treatment. Second, the experimental group was not biased by receiving subjects with a favorable diagnosis or biased IQ testing because both diagnosis and IQ tests were constant across groups. Third, the referral process did not favor the project cases because there were no significant differences between Control Groups 1 and 2 at intake or follow-up, even though Control Group 2 subjects were referred to others by the same agency.

Fourth, subjects stayed within their groups, which preserved the original (unbiased) group assignment.

2. A favorable outcome could have been caused not by the experimental treatment but by the attitudes and expectations of the staff. There are two findings that contradict this possibility of treatment agency (placebo) effects. First, because Control Group 2 subjects had no contact with the project, and because there was no difference between Control Groups 1 and 2 at follow-up, placebo effects appear implausible. Second, the within-subjects study showed that at least one treatment component contributed to the favorable outcome in the intensive treatment (experimental) group.

3. It may be argued that the treatment worked because the subjects were not truly autistic. This is counterindicated by the high reliability of the independent diagnosis and by the outcome data from the control groups, which are consistent with those reported by other investigators (Brown, 1969; DeMeyer et al., 1973; Eisenberg, 1956; Freeman et al., 1985; Havelkova, 1968; Rutter, 1970) for groups of young autistic children diagnosed by a variety of other agencies.

4. The spontaneous recovery rate among very young autistic children is unknown, and without a control group the favorable outcome in the experimental group could have been attributed to spontaneous recovery. However, the poor outcome in the similarly constituted Control Groups 1 and 2 would seem to eliminate spontaneous recovery as a contributing factor to the favorable outcome in the experimental group. The stability of the IQ test scores in the young autistic children examined by Freeman et al. (1985) attests once again to the chronicity of autistic behaviors and serves to further negate the effects of spontaneous recovery.

5. Posttreatment data showed that the effects of treatment (a) were substantial and easily detected, (b) were apparent on comprehensive, objective, and socially meaningful variables (IQ and school placement), and (c) were consistent with a very large body of prior research on the application of learning theory to the treatment and education of developmentally disabled persons and with the very extensive (100-year-old) history of psychology laboratory work on learning processes in man and animals. In short, the favorable outcome reported for the intensive-treatment experimental group can in all likelihood be attributed to treatment.

A number of measurement problems remain to be solved. For example, play, communicative speech, and IQ scores define the characteristics of autistic children and are considered predictors of outcome. Yet the measurement of these variables is no easy task. Consider play. First, play undoubtedly varies with the kinds of toys provided. Second, it is difficult to distinguish low levels of toy play (simple and repetitive play associated with young, normal children) from high levels of self-stimulatory behavior (a psychotic attribute associated with autistic children). Such problems introduce variability that needs immediate attention before research can proceed in a meaningful manner.

The term *normal functioning* has been used to describe children who successfully passed normal first grade and achieved an average IQ on the WISC–R. But questions can be asked about whether these children truly recovered from autism. On the one hand, educational placement is a particularly valuable measure of progress because it is sensitive to both educational accomplishments and social-emotional functions. Also, continual promotion from grade to grade is made not by one particular teacher but by several teachers. School personnel describe these children as indistinguishable from their normal friends. On the other hand, certain residual deficits may remain in the normal functioning group that cannot be detected by teachers and parents and can only be isolated on closer psychological assessment, particularly as these children grow older. Answers to such questions will soon be forthcoming in a more comprehensive follow-up (McEachin, 1987).

Several questions about treatment remain. It is unlikely that a therapist or investigator could replicate our treatment program for the experimental group without prior extensive theoretical and supervised practical experience in one-to-one behavioral treatment with developmentally disabled clients as described here and without demonstrated effectiveness in teaching complex behavioral repertoires as in imitative behavior and abstract language. In the within-subjects studies that were reported, contingent aversives were isolated as one significant variable. It is therefore unlikely that treatment effects could be replicated without this component. Many treatment variables are left unexplored, such as the effect of normal peers. Furthermore, the successful mainstreaming of a 2–4-year-old into a normal preschool group is much easier than the mainstreaming of an older autistic child into the primary grades. This last point underscores the importance of early intervention and places limits on the generalization of our data to older autistic children.

Historically, psychodynamic theory has maintained a strong influence on research and treatment with autistic children, offering some hope for recovery through experiential manipulations. By the mid-1960s, an increasing number of studies reported that psychodynamic practitioners were unable to deliver on that promise (Rimland, 1964). One reaction to those failures was an emphasis on organic theories of autism that offered little or no hope for major improvements through psychological and educational interventions. In a comprehensive review of research on autism, DeMyer et al. (1981) concluded that "[in the past] psychotic children were believed to be *potentially* capable of normal functioning in virtually all areas of development . . . during the decade of the 1970s it was the rare investigator who even gave lip-service to such previously held notions . . . infantile autism is a type of developmental disorder accompanied by severe and, to a large extent, permanent intellectual/behavioral deficits" (p. 432).

The following points can now be made. First, at least two distinctively different groups emerged from the follow-up data in the experimental group. Perhaps this finding implies different etiologies. If so, future theories of autism will have to identify these groups of children. Second, on the basis of testing to date, the recovered children show no permanent intellectual or behavioral deficits and their language appears normal, contrary to the position that many have postulated (Rutter, 1974; Churchill, 1978) but consistent with Kanner's (1943) position that autistic children possess potentially normal or superior intelligence. Third, at intake, all subjects evidenced deficiencies across a wide range of behaviors, and during treatment they showed a broad improvement across all observed behaviors. The kind of (hypothesized) neural damage that mediates a particular kind of behavior, such as language (Rutter, 1974), is not consistent with these data.

Although serious problems remain for exactly defining autism or identifying its etiology, one encouraging conclusion can be stated: Given a group of children who show the kinds of behavioral deficits and excesses evident in our pretreatment measures, such children will continue to manifest similar severe psychological handicaps later in life unless subjected to intensive behavioral treatment that can indeed significantly alter that outcome.

These data promise a major reduction in the emotional hardships of families with autistic children. The treatment procedures described here may also prove equally effective with other childhood disorders, such as childhood schizophrenia. Certain important, practical implications in these findings may also be noted. The treatment schedule of subjects who achieved normal functioning could be reduced from 40 hr per week to infrequent visits even after the first 2 years of treatment. The assignment of one full-time special-education teacher for 2 years would cost an estimated $40,000, in contrast to the nearly $2 million incurred (in direct costs alone) by each client requiring life-long institutionalization.

References

Ackerman, A. B. (1980). *The contribution of punishment to the treatment of preschool aged children.* Unpublished doctoral dissertation, University of California, Los Angeles.

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

Bayley, N. (1955). On the growth of intelligence. *American Psychologist, 10,* 805–818.

Brecht, M. L., & Woodward, J. A. (1984). GANOVA: A univariate/multivariate analysis of variance program for the personal computer. *Educational and Psychological Measurement, 44,* 169–173.

Brown, J. (1969). Adolescent development of children with infantile psychosis. *Seminars in Psychiatry, 1,* 79–89.

Cattell, P. (1960). *The measurement of intelligence of infants and young children.* New York: Psychological Corporation.

Churchill, D. W. (1978). Language: The problem beyond conditioning. In M. Rutter & E. Schopler (Eds.), *Autism: A reappraisal of concepts and treatment* (pp. 71–85). New York: Plenum.

DeMyer, M. K., Barton, S., DeMyer, W. E., Norton, J. A., Allen, J., & Steele, R. (1973). Prognosis in autism: A follow-up study. *Journal of Autism and Childhood Schizophrenia, 3,* 199–246.

DeMyer, M. K., Hingtgen, J. N., & Jackson, R. K. (1981). Infantile autism reviewed: A decade of research. *Schizophrenia Bulletin, 7,* 388–451.

Doll, E. A. (1953). *The measurement of social competence.* Minneapolis, MN: Minneapolis Educational Test Bureau.

Dunn, L. M. (1981). *Peabody Picture Vocabulary Test.* Circle River, MI: American Guidance Service.

Education for All Handicapped Children Act of 1975. Washington, DC: *Congressional Record.*

Eisenberg, L. (1956). The autistic child in adolescence. *American Journal of Psychiatry, 112,* 607–612.

Freeman, B. J., Ritvo, E. R., Needleman, R., & Yokota, A. (1985). The stability of cognitive and linguistic parameters in autism: A 5-year study. *Journal of the American Academy of Child Psychiatry, 24,* 290–311.

Gesell, A. (1949). *Gesell Developmental Schedules.* New York: Psychological Corporation.

Havelkova, M. (1968). Follow-up study of 71 children diagnosed as psychotic in preschool age. *American Journal of Orthopsychiatry, 38,* 846–857.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2,* 217–250.

Leiter, R. G. (1959). Part I of the manual for the 1948 revision of the Leiter International Performance Scale: Evidence of the reliability and validity of the Leiter tests. *Psychology Service Center Journal, 11,* 1–72.

Lotter, V. (1967). Epidemiology of autistic conditions in young children: II. Some characteristics of the parents and children. *Social Psychiatry, 1,* 163–173.

Lovaas, O. I., Ackerman, A. B., Alexander, D., Firestone, P., Perkins, J., & Young, D. (1980). *Teaching developmentally disabled children: The me book.* Austin, TX: Pro-Ed.

Lovaas, O. I., Koegel, R. L., Simmons, J. Q., & Long, J. (1973). Some generalization and follow-up measures on autistic children in behavior therapy. *Journal of Applied Behavior Analysis, 6,* 131–166.

McEachin, J. J. (1987). *Outcome of autistic children receiving intensive behavioral treatment: Residual deficits.* Unpublished doctoral dissertation, University of California, Los Angeles.

McEachin, J. J., & Leaf, R. B. (1984, May). *The role of punishment in motivation of autistic children.* Paper presented at the convention of the Association for Behavior Analysis, Nashville, TN.

Ray, A. A. (1982). *Statistical Analysis System user's guide: Statistics, 1982 edition.* Cary, NC: SAS Institute.

Rimland, B. (1964). *Infantile autism.* New York: Appleton-Century-Crofts.

Rutter, M. (1970). Autistic children: Infancy to adulthood. *Seminars in Psychiatry, 2,* 435–450.

Rutter, M. (1974). The development of infantile autism. *Psychological Medicine, 4,* 147–163.

Stutsman, R. (1948). *Guide for administering the Merrill-Palmer Scale of Mental Tests.* New York: Harcourt, Brace & World.

Thorndike, R. L. (1972). *Manual for Stanford-Binet Intelligence Scale.* Boston: Houghton Mifflin.

Wechsler, D. (1967). *Manual for the Wechsler Pre-School and Primary Scale of Intelligence.* New York: Psychological Corporation.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised.* New York: Psychological Corporation.

Received October 10, 1985
Revision received March 28, 1986 ■