Journal of Autism and Developmental Disorders, Vol. 18, No. 1, 1988

An Evaluation of the Autism Behavior Checklist[1]

Fred R. Volkmar,[2] Domenic V. Cicchetti, Elizabeth Dykens, Sara S. Sparrow, James F. Leckman, and Donald J. Cohen
Yale University

The Autism Behavior Checklist (ABC), an assessment instrument for autistic individuals, was evaluated in a group of 157 subjects, 94 clinically autistic and 63 nonautistic. The two groups differed significantly in ratings of pathology. Both false positive and false negative diagnostic classifications were made when the results of the checklist were compared with clinical diagnosis. Effects of developmental level and age were observed. The ABC appears to have merit as a screening instrument, though results of the checklist alone cannot be taken as establishing a diagnosis of autism. Important issues of reliability and validity remain to be addressed.

Over the course of the past four decades clinicians and investigators have grappled with the problems of diagnosis and definition of the autistic syndrome and its differentiation from other severe developmental problems of childhood. Considerable progress has been made in some areas. It now appears that autism is an early onset disorder characterized by disturbances in social relatedness, communication, and certain information-processing and cognitive abilities (Cohen, Paul, & Volkmar, 1986; Cohen, Volkmar, & Paul, 1986), and, on the basis of several lines of research, it is discontinuous from other severe childhood behavioral disorders (e.g., childhood schizophrenia) and adult psychoses. As Rutter and Garmezy have pointed out (1983, p. 794), infan-

[1]The authors thank Richard Rende and Joseph Coleman for assistance in data analysis, and George M. Anderson for his comments on an earlier version of the manuscript. We thank the subjects, their parents, and their teachers for their cooperation. This work was supported in part by the William T. Grant Foundation, the John Merck Fund, MHCRC Grant 30929, CCRC Grant RR00125, NICHD Grant HD-03008, NIMH Grant MH00418, and Mr. Leonard Berger.

[2]Address all correspondence to Dr. Fred R. Volkmar, Child Study Center, Yale University, P.O. Box 3333, New Haven, Connecticut 06510.

0162-3257/88/0300-0081$06.00/0

tile autism represents, in many ways, the "clearest example of a disease entity in child psychiatry." This consensus is reflected by the inclusion of autism in the official diagnostic and statistical manual (DSM-III; American Psychiatric Association, 1980). As in DSM-III, various attempts to systematize and/or operationalize this diagnostic concept have been made (Ritvo & Freeman, 1978; Rutter, 1978; Cohen, Paul & Volkmar, 1986; however, truly operational definitions have yet to emerge, and problems with the DSM-III scheme have become apparent (Volkmar, Stier, & Cohen, 1985; Volkmar, Cohen, & Paul, 1986).

Another approach to the issue of diagnosis and assessment is exemplified by the various rating scales and diagnostic checklists developed for autism (see Parks, 1983, for a review). These assessment instruments attemp to provide both clinicians and investigators with relatively straightforward measures of current functioning or diagnosis. For example, Rimland (1971) developed a multiple-choice diagnostic checklist in which parents responded to questions related to the child's behavior early in life; this instrument appears to be overly stringent in making the diagnosis of autism (Parks, 1983). Other instruments focus on direct observation of the child, using an assessment instrument, e.g., the Behavior Rating Instrument for Autistic and Atypical Children (BRIAAC; Ruttenberg, Dratman, Franknoi, & Wenar, 1966; Ruttenberg, Kalish, Wenar, & Wolf, 1977) and the Childhood Autism Rating Scale (Schopler, Reichler, DeVellis, & Daly, 1980). Other instruments, such as the Behavior Observation Scale for Autism (Freeman, Schroth, Ritvo, Guthrie, & Wake, 1980), were developed to collect standard observational data for clinical or research purposes. Important questions of the reliability and validity of these scales and checklists remain to be addressed; few studies have systematically examined the reliability and validity of these instruments (Cohen et al., 1978; Parks, 1983).

Various problems are posed for the development of diagnostic rating scales and checklists. First, infantile autism is a relatively heterogeneous syndrome in some respects; e.g., many, though not all, autistic individuals are mentally retarded, and many, though not all, remain mute (Rutter & Garmezy, 1983). Second, considerable change can occur over the course of development and the importance of an "age factor" (Parks, 1983) must be emphasized. Third, assessment instruments have typically employed some variant of a deviance model (Zigler, 1969) in their construction; i.e., deviant or abnormal behaviors are rated and the problem of appropriate comparison groups then arises. Finally, some instruments rely on parental reports, which may be unreliable, particularly in relation to events or behavior occurring some years before completion of the assessment instrument (Robbins, 1963).

The Autism Behavior Checklist (ABC; Krug, Arick & Almond, 1979, 1980) represents the most recent attempt to provide a diagnostic instrument for autism and has already been used for research studies. The ABC is one part of the Autism Screening Instrument for Educational Planning (ASIEP). Other components include standardized observation and description of vocalization, interaction, educational status, and learning rate.

The ABC is completed by teachers and consists of a series of 57 questions, selected from various sources, which are grouped in five areas: sensory, relating, body/object use, language, and social and self-help. The items themselves are dichotomous (i.e. yes/no) but are assigned weights from 1 to 4 on the basis of chi-square analyses of 1,049 completed checklists. The checklist is based on a deviance model, so that higher scores reflect greater deviance or impairment. Some questions relate to historical information, e.g., age at recognition of the syndrome, while others relate to current behavioral functioning. Sums of the weighted scores can be used to operationally differentiate autistic from nonautistic individuals, although the authors also emphasize the importance of careful clinical evaluation to establish a diagnosis of autism. Individuals with a total score of 67 or more (½ standard deviation below the autistic population mean) have a high probability of being autistic, those with scores in the 53-to-67 total score range are questionably autistic, while those with total scores less than 53 are unlikely to be autistic.

Several features of the checklist make it a potentially interesting and attractive assessment instrument. Teachers, rather than parents, are used as raters, though it is suggested that parents may provide valuable assistance and should be used when possible. Administration and scoring are relatively simple and straightforward. The development of the instrument and its standardization are described in some detail. Initial reports of reliability were quite high, though this claim appears to be based on simple correlational methods and percent agreement that do not consider the possible influence of chance agreement. Internal consistency, concurrent validity, and discrimant validity appear to be high. The "age factor" is taken into account by the provision of separate profile charts for different age spans. Comparisons with other groups—e.g., normal, deaf, blind—are provided. In light of their analyses the authors suggest that a high proportion of autistic individuals can be reliably diagnosed with this instrument.

In this paper we report the results of an evaluation of this assessment instrument. ABCs were gathered on a large sample of autistic and a smaller sample of nonautistic but developmentally impaired children and adults from various educational/residential settings. In some cases parents were also asked to complete checklists for purposes of comparison with teacher ratings, and a series of contemporaneous multiple teacher ratings were ob-

tained for evaluation of reliability. In addition, measures of adaptive behavior from the revised Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984a, 1984b) and direct observational data in a smaller number of cases were available for comparison with ABC scores.

METHOD

Subjects

One or more checklists were obtained on 157 individuals, 121 males and 36 females. The mean age of the entire sample was 19.72 years (SD = 12.60). IQ scores from various individually administered tests, including, most commonly the Stanford-Binet, were available for 147 of the subjects; the mean IQ for the sample was 36.80 (SD = 24.30). Th mean mental age of the subjects was 4.12 years (SD = 2.80). The sample included both profoundly retarded individuals (N = 47) and some (N = 14) with IQ scores within the average range. Subjects were selected from several sources, including a university-affiliated school for autistic individuals, a residential facility for the mentally retarded, and a clinic for children with developmental disabilities.

Clinical diagnoses were established using DSM-III criteria prior to scoring and analysis of ABC data. Diagnoses were assigned by experienced clinicians on the basis of clinical assessment and the analysis of available information other than the ABC. It must, of course, be noted that problems with clinical diagnosis in general and the DSM-III scheme in particular (e.g., Volkmar, Cohen, & Paul, 1986) exist and that the use of clinical diagnoses in this study is not meant to imply that a perfect scheme for clinical diagnosis exists. However, clinical diagnoses are obviously of greatest interest in relation to ABC checklist scores since the clinical diagnoses reflect careful consideration of each case on an individual basis by experienced clinicians, and since the ABC checklist was developed using clinical diagnosis. The *clinically autistic group* was composed of those 94 individuals (71 male and 23 female) with clinical diagnosis of infantile autism; their mean age was 17.99 years (SD = 9.11), mean IQ 32.62 (SD = 19.88), and mean MA 4.15 years (SD = 2.78). The *clinically nonautistic* group was composed of the remaining 63 cases (50 male and 13 female) with the following diagnoses: mental retardation (37 cases), atypical pervasive developmental disorder (20 cases), language disorder (4 cases), schizophrenia of childhood onset (2 cases). For the clinically nonautistic group the mean age was 22.30 years (SD = 16.24), mean IQ 43.77 (SD = 29.18), and mean MA 4.07 years (SD = 2.87). The nonautistic group was

older, on average, with a difference approaching statistical significance ($t = 1.92, p < .06$); the two groups did not differ significantly on mental age ($t = .16, p > .80$), though the nonautistic group had a significantly higher IQ ($t = -2.51, p < .05$), reflecting the differences in age but not developmental level between the groups.

Procedure

The standard ABC questionnaire lists every item on a single page along with its weight and symptom area. Since it was conceivable that either the item weightings or groupings might influence responses, a modified version of the checklist was used. This version was identical to the ABC in wording and order of question. It differed only in that the particular weight and component group of each item were *not* indicated and no specific indication that the checklist was designed for autism was made. Each question was answered as either yes or no. No indication was made that the questionnarie was related to the diagnosis of autism as such. Teachers and professional staff were requested to complete an ABC for a subject only if they were well acquainted with the individual. In each case they had known and worked with the individual for a period of at least 1 month and often for several years. Raters were encouraged to answer every item. If more than five items were not responded to, the checklist was discarded from further analysis (28 cases not included in the final data set). If parents were available to the teacher or staff member, with the exception noted below, they were free to request information from the parents. Teachers and parents were, obviously, aware of subjects' past diagnoses. However, they were not aware of the specific purpose of the checklist, and, since the checklist was obtained for children with various developmental problems, it did not appear that raters were aware of the specific purpose for which it had been developed.

In 33 clinically autistic cases parents as well as at least one teacher completed the checklist independently of the parents; i.e., in these instances the teacher was not able to request additional information from parents. In 32 cases (all clinically autistic) two teachers independently completed the checklist. These cases served for an assessment of reliability of the ABC. In instances where more than one checklist was completed, mean scores were calculated for each of the five composite ABC scales and total score.

Data were collected over an 18-month period. While inidvidual checklists were promptly scored and examined, the data were not computer-scored for further statistical analysis until after the completion of the study.

In addition to clinical diagnosis, two other types of data related to validity were available. In 80 cases (35 autistic and 45 nonautistic) recent Vineland Adaptive Behavior Scales (Sparrow et al., 1984a, 1984b) were

available for the subjects. Using a semistructured interview format, the Vineland is administered to a parent or other primary caregiver of the subject and is organized into four domains and several subdomains: Communication (receptive, expressive, and written), Daily Living Skills (personal, domestic, and community), Socialization (interpersonal, play and leisure time, and coping skills), and Motor Skills (gross and fine motor). These four domains (for age 5 and under) or the first three domains (age 6 and above) are then combined to form an Adaptive Behavior Composite Score. These domains can be expressed as raw scores, standard scores, age-equivalent scores, stanines, percentiles, or adaptive levels. In addition, an optional maladaptive domain can be administered to children over 6 years of age. Since maladaptive behaviors are not considered normative at any age, age-equivalent scores are not available. Rather, these scores are either in raw form or in terms of the level of maladaptive behavior (nonsignificant, intermediate, significant). The Vineland is nationally standardized on samples of normal children and adults (birth through 18 years of age) who were matched to the 1980 census for region of country, size of community, race/ethnicity, and level of parental education.

In 41 clinically autistic individuals observational information on behavioral functioning was available. These data were collected using a time sample technique within 1 year of completion of the ABC. Each subject had been observed a minimum of 15 times using a standard coding scheme (see Volkmar, Hoder, & Cohen, 1985, for a description of an earlier version of this coding system). Behaviors coded included frequency of stereotypy, self-abuse, looking at task, staff, or elsewhere, vocalization/communicative acts, echolalia, and correct, partial, or incorrect responses. Data were collected in the familiar school environment by six trained observers. Every fifth observation was done jointly by two observers for calculation of reliability indices. Kappa values for these data ranged from .25 for echolalia to .80 for vocalization/communicative acts. Using the guidelines of Cicchetti and Sparrow (1981), two of the kappas indicated excellent agreement ($> .75$), two good agreement (.60 to .74), and four fair agreement (.40 to .59), with only one kappa (.25) indicating poor agreement after chance correction.

RESULTS

Validity

Discriminant Validity. Scores for each of the five ABC symptom areas and total score were derived for each subject. Means, standard deviations, and associated *t* values for each of the five symptom groups and total score

Table I. ABC Scores for Autistic and Nonautistic Groups[a]

	Clinical diagnosis						
	Autistic		Nonautistic				
Variable	$\bar{X}$	*SD*	$\bar{X}$	*SD*	*t*	*df*	*p*
Sensory	10.20	4.79	6.17	5.19	4.92	126	.001
Relating	19.22	8.12	12.79	8.50	4.73	129	.001
Body/object Use	15.99	7.04	9.52	7.41	5.47	128	.001
Language	11.17	5.11	8.61	5.49	1.80	127	.05
Social/self-help	11.86	4.75	10.69	6.60	2.14	128	.01
ABC total score	68.60	21.99	47.47	20.34	6.18	140	.001

[a]Separate variance estimates used for *t* tests, one-tailed *p* values for ABC scores.

for clinically autistic and nonautistic subjects are presented in Table I. In each of the five symptom areas and for total score these differences achieved statistical significance.

On the basis of the total ABC score, each case was then rated as unlikely to be autistic, questionably autistic, and probably autistic. In instances where multiple ratings (e.g., parents and teacher(s) were available, mean ratings were used. These data, as well as ABC diagnoses for the various nonautistic subjects, are presented in Table II. The distribution of cases within the clinically autistic versus nonautistic and ABC diagnosis differs significantly $\chi^2 = 36.51$, $df = 2$, $p < .001$). Of the clinically autistic subjects, 80.9% were either questionably or probably autistic by ABC total score as compared with 38.1% of nonautistic subjects. Conversely, 19.1% of the clinically autistic subjects and 61.9% of the clinically nonautistic subjects were rated as unlikely to be autistic by ABC total score. An earlier report (Krug et al., 1980) suggested that 86% of clinically autistic subjects received total scores within 1 standard deviation of the ABC autistic standardization sample mean. In this series 78.7% of clinically autistic and 30.2% of clinically nonautistic cases scored within this same range.

Table II. Clinical Diagnosis and ABC Diagnosis

	ABC diagnosis of autism					
	Unlikely		Questionable		Probable	
Clinical diagnosis	*N*	%	*N*	%	*N*	%
Autistic	18	(19.1)	22	(23.4)	54	(57.4)
Nonautistic	39	(61.9)	15	(23.8)	9	(14.3)
Nonautistic diagnostic groups						
Mental retardation	21	(56.8)	10	(27.0)	6	(16.2)
Atypical PDD	12	(69.8)	5	(25.0)	3	(15.0)
Developmental language disorder	4	(100.0)	0	(0)	0	(0)
Childhood schizophrenia	2	(100.0)	0	(0)	0	(0)

If the questionable cases, with intermediate ABC total scores, are ignored, the sensitivity of the checklist is 75% (54/72), the specificity is 81% (39/48), and the overall frequency of correct diagnosis is 77.5% (93/120). The predictive value for a positive diagnosis of the ABC is 86% (54/63) and the predictive value for a negative diagnosis (nonautistic) is 68% (39/57). Substantially the same result is obtained even if cases where teachers and parents made ratings independently are excluded. This is not surprising since final ABC total scores for each subject were based on means of available ratings (parents and teachers) in cases where multiple ratings were available.[3]

A series of *t* tests were used to measure gender, mental age, and placement effects. Females were significantly younger than males ($t = 12.94$, $p < .001$), and a trend for significantly lower MAs in female subjects was also observed ($t = 1.94$, $p < .06$). No significant sex of subjects effects were observed for ABC symptom area scores or for ABC total score, though the comparison for Body/Object Use fell just short of statistical significance ($t = 1.78$, $p = .08$). Clinically autistic subjects enrolled in the intensive educational program (with essentially a 1:1 staff–child ratio) had, as expected, significantly lower IQ scores ($t = 2.80$, $p < .01$) and significantly lower ABC Sensory ($t = 2.91$, $p < .01$) and ABC total scores ($t = 2.32$, $p = .03$).

Concurrent Validity. As described above, two sets of independent measures were available for comparison with ABC scores: data for 80 subjects (35 clinically autistic and 45 nonautistic) from the Vineland and direct observational data for 41 autistic subjects. Correlations of ABC scores and Vineland age-equivalent scores are presented in Table III. Given that the Vineland provides measures of adaptive competence (except for the optional maladaptive behavior domain) while the ABC measures maladaptive behaviors, it was expected that Vineland adaptive behavior scores would correlate negatively with the ABC scores. Similarly, maladaptive behavior domain scores on the Vineland would be expected to correlate positively with ABC scores. An examination of Table III confirms this prediction.

The ABC Total Score and Body-Object Use and Sensory symptom area scores are significantly negatively related to most aspects of adaptive behavior. For the ABC Relating and Social and Self-Help symptom areas

[3]Chi squares were calculated for each of the 57 dichotomous ABC items by clinical group (autistic vs. nonautistic) and were statistically significant in 17 instances (Items: 5,6,12,13,15, 16,27,31,33,35,40,47,50,51,52,53,57,). Seventeen of the individual ABC items exhibited a minimum of 70% overall agreement with clinical diagnosis of autism and chance corrected levels of agreement of .40 or higher (Items: 7,9,10,14,18,22,26,29,31,32,35,39,41,43,48,51, 56).

Table III. Correlations of ABC scores with Vineland[a]

	ABC scores					
	Sensory	Relating	Body-object use	Language	Social/ self-help	ABC total
Vineland age equivalents						
Communication	−26[c]	−04	−35[d]	−04	−16	−37[d]
Receptive	−27[c]	−23[b]	−38[d]	−07	−10	−30[c]
Expressive	−24[b]	−14	−31[c]	−12	06	−18
Written	−20[b]	06	−32[c]	−05	−03	−18
Daily Living	−36[d]	−23[b]	−40[d]	11	−31[c]	−42[d]
Personal	−34[d]	−28[c]	−33[c]	−10	−20[b]	−35[c]
Domestic	−25[b]	−13	−28[c]	−09	−20[b]	−42[d]
Community	−36[d]	−25[b]	−41[d]	−03	−26[c]	−43[d]
Socialization	−28[c]	−30[c]	−34[d]	−08	−07	−37[d]
Interpersonal	−24[b]	−36[d]	−32[c]	−03	−06	−37[d]
Play/leisure time	−26[b]	−24[b]	−32[c]	−05	−17	−42[d]
Coping	−24[b]	−28[c]	−38[d]	−05	−07	−40[d]
Composite	−22[b]	−22[b]	−33[c]	−16	−03	−33[c]
Maladaptive 1	19	19	16	−03	22[b]	22[b]
Maladaptive 1 + 2	24[b]	27[b]	35[c]	03	15	36[c]

[a]Decimal points omitted. $N = 80$ for all comparisons except for maladaptive behaviors, where $n = 62$, except for maladaptive behavior scores correlations calculated using Vineland age-equivalent scores.
[b]$p < .05$, one-tailed.
[c]$p < .01$, one-tailed.
[d]$p < .001$, one-tailed.

significant relationships are confined to predictable Vineland areas, e.g., ABC social and self-help scores with Vineland measures of Daily Living Skills, ABC relating scores with Vineland Socialization scores. Surprisingly, no significant relationships were noted for the ABC language area with any aspect of Vineland performance. A similar pattern of relationships is apparent if only clinically autistic cases are included in the analysis.

The behavioral data for a group of 41 clinically autistic individuals enrolled in an intensive educational program provides another set of measures against which ABC scores can be compared. Since the behavioral data were in the form of proportional scores, Spearman, nonparametric, correlations were calculated for each of the ABC symptom areas and total score (Leach, 1979), and since ABC scores predict greater deviance, one-tailed probability values were calculated for the nonparametric correlations. These results are presented in Table IV. Significant Spearman correlations were observed between ABC Sensory score and Relating score versus echolalia, ABC Body/Object Use score versus stereotypy, looks at task, looks at other, requests to subject, correct and incorrect responses, between ABC Language score versus echolalia and vocalization/communicative acts; and between ABC Social and Self-Help versus looks at other. ABC total score was significantly related to correct and incorrect responses. The pattern of correlations suggests some relationship between ABC scores and behavioral status in relation to particular ABC symptom areas and generally in the expected direction; e.g., maladaptive behaviors (stereotypy, incorrect responses) correlate positively with ABC scores.

Reliability

Interrater Reliability. The original report of interrater reliability (Krug et al., 1979) indicated good (95%) agreement based on 42 independent raters of 14 children. In a subsequent report (Krug et al., 1980) the ABC was used to evaluated 62 clinically autistic individuals and 86% of this group were reported to score within 1 standard deviation of the mean of the standardization sample, though presumably raters were not unaware of diagnosis. Two problems arise in the interpretation of these results. The lack of "blindness" to clinical diagnosis and the intent of the rating instrument may bias estimates of agreement, as may the presence of agreement simply due to chance. The most stringent, and appropriate, measures of reliability include corrections for chance agreement, i.e., that degree of agreement expected by chance alone (see Fleiss, 1981, for a discussion). These issues were addressed in our sample of 32 teacher-teacher ratings of clinically autistic individuals enrolled in an intensive educational program.

Table IV. Correlations of ABC Scores with Behavioral Data[a]

	ABC scores					
	Sensory	Relating	Body-object use	Language	Social/ self-help	ABC total
Behavioral Data						
Stereotypy	12	−05	35[b]	23	20	22
Self-injurious behavior	19	02	06	05	−06	04
Looks at staff	13	01	22	05	−13	04
Looks at task	−11	−13	−28[b]	22	02	−14
Looks at other	10	17	26	14	−31[b]	24
Vocalization/communicative acts	01	−14	−22	37[b]	02	−08
Echolalia	36[b]	27[b]	06	62[c]	08	31[b]
Requests to subject	23	04	−30[b]	06	−18	−04
Correct responses	−09	−22	−31[b]	−01	−19	−27[b]
Partial responses	19	16	20	14	18	−24
Incorrect responses	08	26	41[c]	−14	23	33[b]

[a]Spearman correlations, decimal points omitted. $N = 41$ for all comparisons except for echolalia, where $n = 20$.
[b]$p < .05$, one-tailed.
[c]$p < .01$, one-tailed.

These ratings were made by teachers blind to the purpose of the checklist, to the weightings of items, and to the item groupings. Furthermore, these ratings were made independently by two teachers who had the most experience with each student. Teachers were not able to discuss their ratings with each other *or with parents.* It should be noted that the directions for administration of the ABC suggest that parents should be consulted whenever possible; in this situation such consultation would presumably have inflated estimates of agreement between raters. Further it should be noted, as described previously, that the checklist format was modified to blind raters, insofar as possible, both to the specific purpose of the checklist and to item categories and weights assigned to individual items.

In analyzing the reliability data for individual items the following measures were calculated: the probability of overall agreement (PO), the probability of overall agreement corrected for chance agreement (PC), Kappa, and maximum Kappa (Kmax). Kmax indicates the highest value Kappa could assume given joint agreement on the presence or absence of a trait and was calculated by the formula (POmax-PC)/(1-PC). The ratio of Kappa/Kmax was then evaluated (Cohen, 1960) as an index of interobserver agreement. For 17 of the 57 items the overall probability of agreement was equal to or greater than 70% and the ratio of Kappa/Kmax was greater than or equal to .40 (indicating at least fair chance corrected agreement), e.g., the criteria of Cicchetti and Sparrow (1981) and Fleiss (1981).

Parent Ratings. In 33 clinically autistic cases independent teacher and parent ratings were available. Comparisons between the five symptom areas and ABC total score differed significantly in four comparisons: Relating ($t = 3.03, p < .005$), Body-Object Use ($t = 4.30, p < .001$), Social and Self-help ($t = 4.64, p < .001$) and ABC total score ($t = 4.67, p < .001$). The parents rated significantly more pathology than teachers, suggesting the importance, as stated in the ABC manual, of including parents in the completion of the checklist. Though some ABC items relate to historical information (which parents might presumably be more inclined to report), the strength of the teacher–parent differences suggests that other factors may be operative as well. For example, parents have less opportunity to compare the child with other disordered children and experience the child in the more unstructured home situation.

Internal Consistency, Correlations between the ABC symptoms areas and total scores, age, mental age, and IQ for the entire sample are presented in Table V. Significant correlations were observed between the various ABC symptom areas and ABC total score. Significant negative correlations with age were observed for Body and Object Use, Social and Self-Help, and ABC total score. These results suggest the role of an "age factor" (Parks, 1983), i.e., changes in the expression of symptoms over the course of

Table V. Correlations of ABC Sympton Areas, Age, MA, and IQ[a]

	1	2	3	4	5	6	7	8
ABC scores								
1. Sensory								
2. Relating	54[d]							
3. Body-object use	49[d]	51[d]						
4. Language	43[d]	35[d]	29[d]					
5. Social/self-help	32[d]	47[d]	35[d]	23[c]				
6. ABC total score	74[d]	82[d]	77[d]	58[d]	59[d]			
Other variables								
7. AGE	−06	−14	−26[c]	12	−22[c]	−19[c]		
8. Mental age	−22[c]	−04	−30[d]	−02	05	−16	−02	
9. IQ	−26[c]	−13	−24[c]	−08	03	−20[b]	−47[d]	−67[d]

[a]Decimal points omitted.
[b]$p < .05$.
[c]$p < .01$.
[d]$p < .001$.

development. Significant correlations with IQ and/or MA were observed for Sensory, Body and Object Use, and ABC total score. A principal-component factor analysis of the five ABC symptom area scores, age, and mental age yielded three factors with eigenvalues of greater than 1.00, which accounted for 72% of the total common variance. The first factor (38% of the variance) loaded positively on the five ABC symptom area scores and negatively on age and MA. The second factor (17% of the variance) loaded most strongly on age (−.80) and language (−.48). The third factor (16.1% of the variance) loaded most strongly on mental age (.83) (see Table VI). These results, along with the correlations of symptom areas reported in Table V, suggest that although the ABC symptom areas reflect common elements, they appear to have sufficient variation to justify inclusion as separate scales. They also suggest the importance of both age and mental age for ABC scores.

Split-half reliabilities were computed for the entire sample, for the clinically autistic group, for the checklist as a whole, and for each of the five

Table VI. Factor Analysis—Factor Loadings[a]

Variable	Factor 1	Factor 2	Factor 3
Age	−27	−80	25
Mental age	−23	34	83
ABC sensory score	78	−27	−04
ABC relating	80	06	19
ABC body object use	78	10	−32
ABC language	56	−48	38
ABC social/self-help	61	35	29

[a]Decimal points omitted.

subscales. Split-half reliabilities for the entire sample were .74 for the entire scale, .52 for Sensory, .70 for Relating, .65 for Body/Object Use, .30 for language, and .45 for Social and Self-Help. A similar result was obtained for the autistic group: .70 for the entire scale, .51 for Sensory, .69 for Relating, .59 for Body/Object Use, .29 for language, and .41 for Social and Self-Help.

DISCUSSION

In considering the results it is important to note that the ABC was developed as a screening instrument to identify individuals with high levels of autistic behavior in severely handicapped populations. The advantages of the ABC approach, its ease of administration and scoring, and the complexities of clinical diagnosis have been reviewed earlier in this paper. The attractiveness of such a straightforward assessment instrument is obvious, and ABC total scores have already been used in research studies to indentify individuals with high levels of autistic behavior. As Parks (1983) and others have observed, issues of validity and reliability for all the various assessment and diagnostic instruments for autistic individuals remain to be addressed. In considering the data from this study it is important to note that modifications in the presentation of the checklist were made and, conceivably, may have affected the results. The data from this study suggest some potential usefulness of the ABC as a *screening* instrument as well as its limitations as a *diagnostic* instrument. These limitations are particularly important if this instrument is used in an attempt to simplify the diagnosis of autism for research purposes.

While the total ABC score provides an index of high levels of maladaptive or deviant behaviors commonly observed in autistic individuals, it should be noted that when the recommended scoring procedures are employed, significant numbers of individuals are misclassified. Almost 20% of clinically diagnosed individuals were so misclassified as probably not autistic. False negatives may be more common in higher-functioning autistic individuals; of the five clinically autistic subjects with IQ's greater than 70, three were classified as nonautistic by ABC total score, one as possibly autistic, and one as autistic. Although based on a small number of cases, this suggests that diagnostic validity may be more questionable with higher intellectually functioning subjects. Similarly, nearly 40% of clinically nonautistic individuals are classified as either possibly or probably autistic, reflecting the high levels of autistic behaviors in a nonautistic, developmentally disabled sample. ABC symptom area scores, with the exception of language scores, generally relate in predictable ways to

scores from the Vineland, a well-normatized assessment instrument of adaptive behaviors. The ABC language scale was not related to any aspect of adaptive behavior and had the lowest internal consistency, though it was related to data derived from direct observation.

Interrater reliabilities, using chance corrected and overall percent agreement measures, were strong for only a minority of items. Previous reports of reliability (Krug et al., 1979, 1980) have employed simple percent agreement or correlational methods that do not incorporate corrections for chance agreement. Although the particular format of presentation employed in this study may conceiveably have reduced both the overall ratings of severity and interobserver agreement, there appears to be cause for concern about the reliability of the instrument. Parents appear to rate more pathology than teachers, and, as implied in the directions for administration of the ABC, it appears to be important to use parents in the completion of the checklist. The use of such a procedure would have probably have increased measures of teacher reliability. It is also likely that the specific identification of the instrument as a checklist for autistic individuals and indication of item weightings might have increased teacher ratings of pathology.

Correlational and factor-analytic methods suggest the importance of both developmental level and age on ratings using the checklist. Some aspects of behavior clearly change over the course of an autistic individual (Lotter, 1978). The emergence of an "age factor" is not unexpected (Parks, 1983), and the ABC makes some provision for the effect of age by including profiles for different chronological age ranges. Indeed, the authors suggest that repeated administration of the checklist may be helpful in monitoring an individual child, implying that some change is expected. Further, the authors note the importance of obtaining additional information in regard to appropriate educational placement.

The present data suggest that the ABC may have its greatest usefulness as a screening instrument, i.e., to identify individuals for whom there is some question of the diagnosis of autism. However, the data also suggest that ABC scores alone cannot substitute for careful clinical assessment and diagnostic evaluation and that research studies should not simply equate ABC scores with diagnosis. Some reports (e.g., Hagerman, et al., 1986) have used the ABC to identify individuals with high levels of "autistic behaviors" as defined by the ABC and suggest on this basis that certain diagnostic groups—e.g., fragile X symdrome—exhibit marked increases in autism. Such an approach appears to be overly simplistic. It is not sufficient to equate completion of any checklist or rating instrument with clinical diagnosis, though provision of any detailed information regarding diagnosis and assessment is praiseworthy. Additional research may clarify

issues of reliability and validity of the ABC. Indeed, independent studies of reliability and validity of all the various assessment and diagnostic instruments for this population are sorely needed. Comparisons between various assessment instruments, both those developed specifically for autism and those developed for the normal population (Volkmar, Sparrow, et al., 1987; Fein, Pennington, Markowitz, Braverman, & Waterhouse, 1986), may be helpful in this regard.

REFERENCES

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

Cicchetti, D. V., & Sparrow, S. S. (1981). Developing criteria for establishing interrater reliability of specific items in a given inventory. *American Journal of Mental Deficiency, 86,* 127-137.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurements, 20,* 37-46.

Cohen, D. J., Caparulo, B., Gold, J., Waldo, M., Shaywitz, B., Ruttenberg, B., & Rimland, B. (1978). Agreement in diagnosis: Clinical assessment and behavior rating scales for pervasively disturbed children. *Journal of the American Academy of Child Psychiatry, 17,* 689-695.

Cohen, D. J., Paul, R., & Volkmar, F. R. (1986) Issues in the classification of pervasive and other developmental disorders: Towards DSM IV. *Journal of the American Academy of Child Psychiatry, 25,* 213-220.

Cohen, D. J., Volkmar, F. R., & Paul, D. (1986) Issues in the classification of "pervasive developmental disorders": History and current status of nosology. *Journal of the American Academy of Child Psychiatry, 25,* 158-161.

Fein, D., Pennington, B., Markowitz, P., Braverman, M., & Waterhouse, L. (1986). Towards a neuropsychological model of infantile autism: Are the social deficits primary? *Journal of the American Academy of Child Psychiatry, 25,* 198-212.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.), New York: Wiley.

Freeman, B. J., Schroth, P., Ritvo, E., Guthrie, D., & Wake, J. (1980). The Behavior Observation Scale for Autism (BOS): Initial results of factor analyses. *Journal of Autism and Developmental Disorders, 10,* 343-346.

Hagerman, R. J., Jackson, A. W., Levits, A., et al. (1986). An analysis of autism in 50 males with the fragile X syndrome. *American Journal of Medical Genetics, 23,* 359-374.

Krug, D. A., Arick, J. R., & Almond, P. G. (1979) Autism screening instrument for educational planning background and development. In J. Gilliam (Ed.), Autism: Diagnosis, instruction, management and research. Austin: University of Texas at Austin Press.

Krug, D. A., Arick, J. R., & Almond, P. G. (1980). Behavior checklist for identifying severely handcapped individuals with high levels of autistic behavior. *Journal of Child Psychology and Psychiatry, 21,* 221-229.

Leach, C. (1979). *Introduction to statistics: A nonparametric approach for the social sciences.* New York: Wiley.

Lotter, V. (1978). Follow-up studies. In M. Rutter & E. Schopler (Eds.), *Autism: a reappraisal of concepts and treatment.* New York: Plenum.

Parks, S. L. (1983). The assessment of autistic children: A selective review of available instruments. *Journal of Autism and Developmental Disorders, 13,* 255-267.

Rimland, B. (1971). The differentiation of childhood psychoses: An analysis of checklists for 2,218 psychotic children. *Journal of Autism and Childhood Schizophrenia, 1,* 161-174.

Ritvo, E. R., & Freeman, B. J. (1978). National Society for Autistic Children definition of the syndrome of autism. *Journal of Autism and Childhood Schizophrenia, 8,* 162-169.

Robbins, L. C. (1963). The accuracy of parental recall of aspects of child development and of child rearing practices. *Journal of Abnormal and Social Psychology, 66,* 261-270.

Ruttenberg, B. A., Dratman, M. L., Fraknoi, J., & Wenar, C. (1966). An instrument for evaluating autistic children. *Journal of the American Academy of Child Psychiatry, 5,* 453-478.

Ruttenberg, B. A., Kalish, B. I., Wenar, c., & Wolf, E. G. (1977). *Behavior rating instrument for autistic and other atypical children* (rev. Ed). Philadelphia: Developmental Center for Autistic Children.

Rutter, M. (1978). Diagnosis and definition. In M. Rutter & E. Schopler (Eds.), *Autism: A reappraisal of concepts and treatment.* New York: Plenum.

Rutter, M., & Garmezy, N. (1983). Developmental psychopathology. In E. M. Hetherington (Ed.), *Handbook of child psychology (4th ed.) Volume IV: Socialization, Personality, and Social Developmental* (pp. 775-911). New York: Wiley.

Schopler, E., Reichler, R. L., DeVellis, R. F., & Daly, K. (1980), Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journals of Autism and Developmental Disorders, 10,* 91-103.

Sparrow, S., Balla, D., & Cicchetti, D. (1984a). *Vineland Adaptive Behavior Scales (Survey Form).* Circle Pines, Minnesota: American Guidance Service.

Sparrow, S., Balla, D., & Ciccheti, D. (1984b). *Vineland Adaptive Behavior Scales (Expanded Form).* Circle Pines, Minnesota: American Guidance Service.

Volkmar, F., Cohen, D. J., & Paul, R. (1986). An evaluation of DSM-III criteria for infantile autism. *Journal of the American Academy of Child Psychiatry, 25,* 190-197.

Volkmar, F. R., Hoder, E. L., & Cohen, D. J. (1985). Compliance, 'negativism,' and the effects of treatment structure in autism: A naturalistic, behavioral study. *Journal of Child Psychology and Psychiatry, 26,* 865-877.

Volkmar, F., Sparrow, S. A., Goudreau, D., Cicchetti, D. V., Paul, R., & Cohen D. J. (1987). Social Deficits in autism: An operational approach using the Vineland Adaptive Behavior Scales. *Journal of the American Academy of Child Psychiatry, 26,* 156-161.

Volkmar, F. R., Stier, D. M., & Cohen, D. J. (1985). Age of onset of pervasive developmental disorder. *American Journal of Psychiatry, 142,* 1450-1452.

Zigler, E. (1969). Developmental versus difference theories of mental retardation and the problem of motivation. *American Journal of Mental Deficiency, 73,* 536-549.