Journal of Autism and Developmental Disorders, Vol. 24, No. 5, 1994

Autism Diagnostic Interview–Revised: A Revised Version of a Diagnostic Interview for Caregivers of Individuals with Possible Pervasive Developmental Disorders[1]

Catherine Lord[2]
The University of Chicago

Michael Rutter
Institute of Psychiatry

Ann Le Couteur
Royal Free Hospital

Describes the Autism Diagnostic Interview–Revised (ADI-R), a revision of the Autism Diagnostic Interview, a semistructured, investigator-based interview for caregivers of children and adults for whom autism or pervasive developmental disorders is a possible diagnosis. The revised interview has been reorganized, shortened, modified to be appropriate for children with mental ages from about 18 months into adulthood and linked to ICD-10 and DSM-IV criteria. Psychometric data are presented for a sample of preschool children.

[1]We acknowledge the help of all the parents who participated in these interviews and of Sharon Storoschuk, Joyce Magill, Cathy Mulloy, and Lyn Anderson-Cook who conducted interviews. Support was provided by Public Health Service Grant #7 R01 MH46865-03, Training in RESH Diagnoses of Autism and Spectrum Disorder 7-T35-MH 19726-02 and the Alberta Heritage Fund for Medical Research, first author and the MRC Child Psychiatry Unit.

[2]Address all correspondence to Catherine Lord, Department of Psychiatry, University of Chicago, 5841 S. Maryland Avenue, MC 3077, Chicago, Illinois 60637. Data for all items, including those not in the algorithm are available upon request.

0162-3257/94/1000-0659$07.00/0

The Autism Diagnostic Interview–Revised (ADI-R) is a newly modified version of the Autism Diagnostic Interview (ADI; Le Couteur et al., 1989), a standardized, semistructured, investigator-based interview for caregivers of autistic individuals, which provides a diagnostic algorithm for the ICD-10 definition of autism (World Health Organization [WHO], 1992) and DSM-IV (American Psychiatric Association [APA], 1993). In this report, we describe the revision of the ADI and procedures for training and use of the new instrument, the ADI-R.

Psychometric properties for the original ADI were provided for a carefully selected, blindly interviewed and coded, sample of 16 autistic and 16 mentally handicapped children and adults covering a range of IQs and chronological ages (Le Couteur et al., 1989). The interview yielded separate scores in the areas of communication, social skills, and restricted, repetitive and stereotyped behaviors, as well as early history, allowing the separate quantification of severity in each of these domains. Since work on the ADI began in the early 1980s, a number of developments have suggested revisions that allow items to more accurately reflect autism-specific deficits and current theoretical perspectives (see Lord & Rutter, 1994). The original ADI was intended for research purposes and provided behavioral assessment for subjects with a chronological age of 5 years or above and a mental age of at least 2 years. However, because most autistic children are now diagnosed during the preschool years (Short & Schopler, 1988), there was a need for an instrument that could differentiate autism from other disorders as they present in very young children. In addition, if the interview were to be used regularly for clinical purposes, it was important to increase its efficiency and shorten its length, so that it could be incorporated more easily into existing multidisciplinary diagnostic assessments.

CHANGES IN ORGANIZATION

The original ADI began with a series of opening questions; followed by questions concerning infancy and early (under age 5 years) development in social skills, communication and play; followed by an overlapping, but not identical, set of items convening social and communication skills at the time of interview; a section on restricted and repetitive behaviors and interests scored for both current behavior and their occurrence "ever"; and a final section on general behavior difficulties. Because two goals of the revision were to make the interview briefer and more appropriate for younger children, questions concerning early development were consolidated with those concerning careened behaviors. Thus, the ADI-R consists

of five sections: opening questions; questions on communication (both early and current); those on social development and play (again both early and current); enquiries about repetitive and restricted behaviors (all scored for both current and ever judgments); and a reduced number of questions concerning general behavior problems. It is now possible, after substantial practice, to give the interview to the parent of a 3- or 4-year-old suspected of autism in approximately 1½ hours; for older children, the interview may take somewhat longer.

Modification of Items

Research in the last 10 years has provided much useful information clarifying the nature of autism-specific deficits in social and communication skills (Baron-Cohen, Tager-Flusberg, & Cohen, 1993; Lord & Rutter, 1994). Better discrimination between mentally handicapped autistic children and nonautistic severely mentally handicapped children is possible through more focused descriptions of contexts in which normally developing and mentally handicapped children behave in consistent ways from early in development. However, it is important to acknowledge that there is difficulty differentiating autism when it is accompanied by *profound* mental retardation; this difficulty applies to all diagnostic instruments and not just the ADI-R. Revisions in the area of communication were aimed at identifying autism-specific aspects of each behavior. For example, an item on pointing in the original ADI was modified so that the behavior of interest is now pointing at a distance to express interest, thus excluding pointing as an instrumental gesture. Several other items were modified to become more general in focus. A question concerning sensitivity to noise was broadened to include sensitivity to most or all loud noises; in the ADI-R, idiosyncratic sensitivity to highly specific, not necessarily loud, noises is now scored as part of another item concerning abnormal, idiosyncratic negative reactions to any kind of sensory stimuli. Changes in social items occurred along the same lines. The quality of social overtures is now scored with particular reference to the child's attempts to secure help; other items code the extent to which the child makes social overtures and the range of intentions manifested. The emphasis for social smiling was shifted from the age of first social smile to whether and if so, when, the child smiles reciprocally with others. Range of facial expression is scored only during communicative use.

Repetitive use of objects, such as wheel spinning, is now distinguished from compulsions or rituals that involve an end point and a sequence of actions toward that end. Difficulty with changes in the subject's own routines,

such as that of a child who is upset if he does not drink from the same cup each night, is now scored separately from resistance to minor changes in the environment, such as seen in a child who becomes distressed if the furniture in his parents' bedroom is moved. Because of the addition of these new items, the scoring in this area has also been changed in the algorithm. Some pairs of items that fall under similar ICD-10/DSM-IV headings (e.g., stereotyped and repetitive motor mannerisms) are grouped together and the highest score from either item within the group is used in the algorithm. The purpose of this approach is to allow finer distinctions among behaviors, without giving undue credit to possibly overlapping behaviors in the algorithm.

Additional Items in the ADI-R

Several new social and communication items were added to identify social behaviors expected in normally developing or nonautistic mentally handicapped children under age 5 that would be abnormal or absent in young autistic children and items expected to be associated with autism in young children. These behaviors include use of another's body to communicate, showing and directing attention, and interest in and response to other children. Items were also added to provide better coverage of abnormalities as they appear in older high-functioning children and adults. For example, there is a question about talk expressing interest in others and a question about relatively appropriate, but unusually intense, circumscribed interests. An item concerning behaviors associated with social avoidance related to anxiety was added in order to tap behaviors that differentiate children and adolescents with fragile X from those with autism (Wolff, Gardner, Paccia, & Lappen, 1989). Similarly, there are questions about midline hand-wringing movements and about hyperventilation in order to cover behavior that is characteristic of Rett syndrome (Hagberg, Aicardi, Dias, & Rasmos, 1983; Olson & Rett, 1991; Tsai, 1992). More specific questions about the age when abnormalities were first manifest were added, as well as questions about loss of skills and progressive deterioration, again aimed at providing more accurate information for differential diagnosis between autism and syndromes such as Rett syndrome or disintegrative disorders.

Elimination of Items from the Original ADI

Items judged to be redundant, limited in reliability, or applicable to only a small proportion of children with pervasive developmental disorders

(PDD) were eliminated. Such items included questions about sense of humor, pica, sharing activities, and understanding plots and instructions.

Definition of Items

Definitions of all behavioral items have now been provided in the ADI-R; these complement and clarify the details given in each coding. It is the interviewer's responsibility to determine whether, for each item, the behavior described truly meets the specified criteria.

Administration of Interview

The interview continues to focus on caregiver descriptions of actual behavior as it has occurred in the subject's daily life. Scoring is made on the basis of the interviewer's judgment of the code that best fits the behaviors described by the caregiver, rather than on judgments made by the informant. Although the administration of the ADI-R requires a substantial amount of time, most parents and caregivers find it a relatively comfortable experience, because they are allowed to describe important aspects of their child's behavior in their own words. Parents often report that the process left them with a sense that the interviewer valued their impressions and opinions and wanted to know more about their child than could be determined by observation in a clinic. The experience of the interview, as part of a multidisciplinary intake assessment, also helps some parents have a better understanding of the factors, particularly social behaviors, that are being evaluated in order to reach a diagnosis. Questions in the interview are ordered by content area, but within these, they are deliberately written and sequenced to provide caregivers with opportunities to describe positive aspects of their child's behavior and development and try to minimize the effect of having repeatedly to answer questions in a negative way.

Scoring and Interpretation

Coding methods remain the same as the ADI, with most items coded *no definite behavior of the type specified* (0), *behavior of the type specified probably present but defining criteria not fully met* (1), and *definite abnormal behavior of the type described in the definition and coding* (2), with a code of 3 used occasionally to indicate extreme severity. Each item is scored for current behavior, with the exception of a few

items where the behavior is relevant only during particular age periods. For these items, specific age restrictions are given. For example, *imaginative play, imaginative play with peers,* and *group play* are coded only between ages 4 and 10 years, and *reciprocal friendships* only above 10 years of age. *Circumscribed interests* is only scored for children 10 years and older. These items were omitted from the present analyses because of the group age of the sample. Each item describing the presence of a qualitatively abnormal behavior (i.e., a positive item) is also scored as to whether the behavior "ever" occurred for a period lasting as long as 3 months. For these items, the "ever" coding is a lifetime measure that includes the current period. Such behaviors can be coded for anytime after the subject has achieved a mental age greater than 18 months.

Each item inquiring about the lack of a behavior or skill associated with normal development (i.e., a negative item), is coded for its most abnormal manifestation between the ages of 4 and 5 years (i.e., 48–60 months), in addition to the current situation at the time of interview. The rationale of the focus on the 4- to 5-year age period concerns the advantages of a standard age period for comparative purposes, together with the desirability of selecting an age that is high enough to provide an adequate range of behavior and low enough to precede possible major changes in behavior.

Algorithm

An algorithm for diagnosis comparable to that used in the original ADI (Le Couteur et al., 1989) was generated by selecting ADI-R items that most closely depicted the specific abnormalities described in the clinical descriptions and diagnostic guidelines from the DSM-IV and ICD-10 (Volkmar et al., 1993; WHO, 1992). As shown in Table III, using this procedure for lifetime diagnosis for subjects over the age of 4, 16 items were identified to measure reciprocal social interaction, 13 to measure communication, and 8 to measure restricted, repetitive behaviors. Within these areas, scores are also computed for each subdomain of abnormality as listed in the diagnostic criteria for research (WHO, 1992), as shown in the tables. The scores are transferred directly from ADI-R items but, in order to ensure that undue weight is not placed on individual items, severity codes of 2 and 3 are both treated as if they were 2. Algorithm cutoffs were determined by generating ROC curves with the present data and identifying the point within each area that yielded the best combination of sensitivity and

specificity with both exceeding .90. The intention is to use one algorithm for children from mental ages of 18 months through adulthood, with three versions containing minor modifications: (a) a lifetime version; (b) a version based only on current behavior; and (c) a version for use with children under the age of 4 years. The last is necessary because obviously all items based on behavior above this age are inapplicable. Data collected on the earlier version of the ADI as part of the DSM-IV field trials indicated relatively little differences between the current and lifetime versions, with the greatest discrepancies emerging for high-functioning adults (Lord et al., in press). Use of the ADI-R with very young preschool children is the topic of a separate paper (Lord, Storoschuk, Rutter, & Pickles, 1993). Psychometric data presented here are for a sample of preschool children with mental ages greater than 18 months, because of earlier findings that the standard algorithm did not discriminate autistic children with mental ages below 18 months from very young children with severe mental handicaps (Lord et al., 1993). This sample was selected because children are now most commonly diagnosed during preschool years and ensuring adequate reliability and validity for them seemed an appropriate place to start.

The algorithm, based on ICD-10/DSM-IV guidelines, specifies a cutoff score of 8 on communication items for verbal subjects (who, as in the original ADI, were operationally defined as individuals scoring 0 on the "level of language" item, indicating use of three-word phrases, spontaneous or echoed, that sometimes contain a verb) and a cutoff of 7 for nonverbal subjects. For all subjects, a minimum score of 10 on social items and 3 for restricted and repetitive behaviors was identified. Because the ADI-R provides much more detailed information on a wide range of specific behaviors than that ordinarily available, we have followed a total score approach. This follows the same principles as the official ICD-10 research criteria (WHO, 1993) in requiring abnormalities in all domains, but avoids the dilemma of deciding how many individual items are required for each "symptom" to be present. This is advantageous because the separation of individual symptoms within each domain is necessarily somewhat arbitrary in view of the fact that all are thought to reflect the same basic deficits. The total score approach also reduces error stemming from the inevitable unreliability of single symptom judgments. To meet ICD-10/DSM-IV draft diagnostic criteria for autism, an individual must meet criteria in each of the three content areas in the guidelines for research diagnoses, as well as exhibiting some abnormality in at least one area by 36 months of age, as described by the caregiver or judged by the interviewer.

PSYCHOMETRIC DATA: STUDY I: RELIABILITY

Method

Subjects

Ten autistic (8 male, 2 female) and 10 mentally handicapped or language-impaired (8 male, 2 female) children served as subjects for initial analyses concerning the reliability of the ADI-R for preschool children. Autistic children were selected from the Communication and Behavior Disorder and Developmental Pediatrics or Clinics at Glenrose Hospital in Edmonton, Alberta, Canada. All autistic children had received independent clinical diagnoses based on DSM-III (APA, 1980) and ICD-10 draft criteria (WHO, 1987) by a clinical psychologist and child psychiatrist within 6 months of participating in the study. Mentally handicapped and language-impaired children were recruited from the same clinical and local noncategorical preschool programs for mentally handicapped children. None had ever received diagnoses of autism or PDD according to teacher or parent reports or hospital records. Although all of the nonautistic children attended a preschool program for mentally handicapped children, two tested as having nonverbal IQs close to or greater than 100, but had significant expressive and receptive language delays. Children from both groups were all living with at least one biological parent; parents were all native English speakers with at least a Grade 11 education. All children in both diagnostic groups had shown significant language delays prior to 36 months of age.

Children who were nonambulatory or had other marked motor impairments, had other than mild, remediable sensory impairments or identifiable syndromes (e.g., Down syndrome, Rett syndrome) or were judged by their teachers to be functioning below the 12-month level overall were excluded from the study because of expected differences in parental perceptions and difficulty in separating deficits associated with autism from sensory or motor impairments or profound mental retardation (Lord et al., in press). Ethnic representation was 15% Asian, 10% West Indian, and 75% white.

All children were assessed using the Merrill-Palmer Scale of Mental Tests (Stutsman, 1931), scoring only nonverbal items. Because the autistic children typically had difficulty with several of the imitation items in the first level of the test, a basal score on the Merrill-Palmer was operationally defined as passing five performance items at the 18- to 23-month level or five performance items at the 24- to 29-month level. Two autistic children and one mentally handicapped child did not reach either criterion and so were administered the

Bayley Scales of Mental Development (Bayley, 1969). To be consistent across tests, all scores reported here are ratio IQs.

The two groups, autistic and nonautistic mentally handicapped/language impaired (MH/LI), did not differ in chronological age, with age ranges in both cases from 36–59 months and mean ages of 48.9 (SD = 12.2) and 50.1 (SD = 15.7) months, respectively. Mean nonverbal IQ/DQ was 64.12 (SD = 32.86) for the autistic children compared to 63.80 (SD = 23.61) months for the nonautistic children. Mental ages ranged from 21–74 months.

Procedures

Videotaped interviews with mothers of children described above were conducted by one of three experienced interviewers, unfamiliar with the children and blind to their diagnoses. The interview consisted of the ADI-R with the following modifications: All items were administered for current behavior, with "current" defined as a behavior that had occurred on a regular basis over the course of at least 1 month in the last 3 months. In addition, a number of items were asked retrospectively for specific ages, usually 2 years (i.e., 24 months) or infancy (defined as "ever" occurring between 6–18 months chronological age). Interviews lasted between 60 and 90 minutes. Each videotaped interview was independently scored by four medical or graduate students also blind to diagnosis. Prior to the scoring, coders had spent 6 weeks watching and scoring videotapes of the ADI-R, giving practice ADI-Rs with supervision, observing assessments, treatment sessions and groups for autistic children and adults, and visiting schools and day programs for individuals with autism and other developmental and psychiatric disorders. Before reliability coding began, each coder had achieved reliabilities of greater than 90% on three consecutive scorings of live or videotaped ADI-Rs with consensus codes determined by the first author and at least one other coder.

Data Analysis

Reliability was calculated using percentage exact agreement and weighted kappa (Cohen, 1968) for each of the six rater pairs. Multirater weighted kappas (Conger, 1980), reported in the following tables were computed across raters, as was mean percentage agreement. Items appropriate only for children over age 4 years were not analyzed. Data for items not

Table I. Interrater Reliability for ADI-R Algorithm Items from ICD-10/DSM-IV[a]

	κ_w (n = 20)	Agreement
Qualitative Abnormalities in Reciprocal Social Interaction		
B1. Failure to use eye-to-eye gaze, facial expression, body posture and gesture to regulate social interaction (.93)		
Direct gaze	.70	.91
Social smiling	.64	.90
Range of facial expression used to communicate	.86	.96
B2. Failure to develop peer relationships (.93)		
Imaginative play with peers[b]	.89	.95
Interest in children	.69	.90
Response to children	.72	.90
Group play with peers[b]	.88	.96
B3. Lack of social-emotional reciprocity and modulation to context (.96)		
Use of other's body	.84	.93
Inappropriate facial expressions	.67	.92
Quality of social overtures	.71	.91
Appropriateness of social responses	.71	.95
Offers comfort	.87	.95
B4. Seeking to share own enjoyment (.95)		
Showing and directing attention	.70	.91
Offering to share	.85	.94
Seeking to share in own enjoyment	.80	.93
Qualitative Impairments Communication and Language		
C1. Delay or total lack of spoken language, not compensated by gesture (.95)		
Pointing to express interest	.80	.92
Conventional gestures	.74	.88
Nodding head	.86	.94
Head shaking	.89	.96

C2V. Relative failure to initiate or sustain conversational interchange (.94)[b]		
Social chat	.77	.93
Reciprocal conversation	.84	.93
C3V. Stereotyped and repetitive use of language (.94)[b]		
Stereotyped utterances	.85	.95
Inappropriate questions	.70	.91
Pronominal reversal	.76	.92
Neologisms/idiosyncratic language	.73	.95
C4. Lack of varied spontaneous make-believe or social imitative play (.94)		
Spontaneous imitation	.69	.90
Imaginative play	.85	.93
Imitative social play	.69	.90
Restricted, Repetitive Behaviors and Interests		
D1. Encompassing preoccupations (.93)		
Unusual preoccupations	.64	.90
D2. Apparently compulsive adherence to nonfunctional rituals (.94)		
Verbal rituals	.83	.93
Compulsions/rituals	.63	.90
D3. Stereotyped and repetitive motor mannerisms (.94)		
Hand and finger mannerisms	.69	.90
Other complex mannerisms	.86	.94
D4. Preoccupation with part-objects or nonfunctional elements of materials (.95)		
Repetitive use of objects	.67	.91
Unusual sensory interests	.82	.95

[a] Intraclass correlations for each area are presented immediately after the area title. Intraclass correlation for all nonverbal communication items (.94); Intraclass correlation for all verbal communication items (.94); Intraclass correlation for social total (.97).
[b] $n = 11$.

included in the final version of the ADI-R algorithm are available from the authors upon request.

Results

Table I provides reliability data for algorithm items concerned with reciprocal social interaction; the remaining social items are reported in the Appendix. Multirater weighted kappa levels exceeded .70 for 12 out of 15 social algorithm items. Weighted kappas for all individual rater pairs exceeded .63. Lower kappas generally reflected limited variability, either with relatively few scores of 2 (e.g., *inappropriate facial expression,* for which few children scored as markedly abnormal) or with few scores of 0, indicating no deficit (e.g., *social smiling,* for which even the language impaired/mentally handicapped children rarely received scores of 0). Nonalgorithm items also showed consistently high interrater reliability with the exception of *arms up to be lifted, discrimination of parents,* and *attention to voice* for which weighed kappas were low (.52, .54 and .59, respectively) because of low variability, though percentage agreement was high (90, 96, and 87, respectively).

Items concerning verbal skills, particularly those describing language abnormalities, were all nearly identical in content to those in the original ADI and were scored for only the 11 verbal subjects. However, there were many new or modified items describing nonverbal aspects of communication. These were scored for all 20 reliability subjects. Multirater weighted kappas equaled or exceeded .69 for all 13 communication items. Nonalgorithm communication items were also generally reliable.

Restricted and Repetitive Behaviors

Weighted kappa levels equaled or exceeded .63 for all 7 items in the area of restricted and repetitive behavior (*circumscribed interests* was not scored); mean percentage agreement was above 90 for all items. Weighted kappas for individual rater pairs were consistently above .55. Kappas were lower than other domains because of more frequent scores of 0 that had occurred for social or communication items.

Reliability for Rater Pairs

For the six rater pairs, mean weighted kappas across all items ranged from .73 to .78, with mean weighted kappas for algorithm items exceeding .75 for each pair. Mean percentage agreement across all

items ranged from 90 to 93 for each pair, with agreement for algorithm items exceeding 92 for all pairs. As shown in Tables I and III, for algorithm items, intraclass correlations for domain scores ranged from .93 to .97.

Internal Consistency

Cronbach's alpha was run for each domain separately to assess the internal consistency of each area. For the 15 items in the social area, item–total correlations ranged from .54 (for *direct gaze*) to .77 (*quality of social overtures*), with an alpha of .95. For restricted and repetitive behaviors, item–total correlations ranged from .30 (*compulsions and rituals*) to .53 (*unusual sensory behaviors*), with an alpha of .69. For the 11 verbal subjects (those who had three-word phrases), item–total correlations ranged from –.06 (*inappropriate questions*) to .77 (*instrumental gestures*) with an alpha of .85. For all subjects, item–total correlations for communication items ranged from .45 (*imitative social play*) to .70 (*conventional, instrumental gestures*) with an alpha of .84.

Reliability Over Time

Six mothers (4 of autistic children, 2 of mentally handicapped children) were reinterviewed 2–3 months later by a different interviewer unfamiliar with the child and the previous interview. Although the small samples limit the interpretability of statistics, exact agreement exceeded 83% ($K_w > .55$) for all but 6 items, with a mean of 91% ($M\ K_w = .72$). On these 6 items, exact agreement occurred in 4 out of 6 rater pairs.

STUDY 2: VALIDITY

Method

Subjects

After the reliability study was completed, an additional 30 subjects were recruited for the validity analyses (15 autistic, 15 nonautistic), resulting in a total of 25 subjects in each group. The same criteria for inclusion in the study for autism and for mental handicap/language impairment were

used as described earlier. Additional subjects were recruited from the sources in Edmonton, Alberta, described previously, from the Greensboro-High Point TEACCH clinic in North Carolina and from local noncategorical programs for language-impaired and mentally handicapped children in Greensboro and the surrounding area.

Children were recruited into two groups of equivalent chronological age (CA) in months: autistic (*M* CA = 46.76, *SD* = 10.73) and nonautistic (*M* CA = 44.72, *SD* = 13.74), shown in Table V. Ethnic distribution was 12% African American and West Indian, 82% white, 6% Asian, Hispanic, and Native American. It was equivalent across language and diagnostic groupings. Social class, measured by occupational status of father (Treiman, 1977) covered a large range, with mean values for each diagnosis ranging from 43.69–52.14; all in the middle class. Additional subjects were assessed using the Merrill-Palmer and the Bayley, as described for the reliability study. Mean IQ/DQS were 71.88 (*SD* = 21.33) for the autistic children and 71.48 (*SD* = 20.09) for the nonautistic children, with mental ages of 34 and 32 months, respectively.

Procedures

Interview procedures were similar to those in the reliability study except that coding for the additional subjects was made on the basis of live interviews of mothers or both parents together. Three additional interviewers were added to the research team over the course of data collection. Prior to data collection, each accompanied the experienced research associates on interviews, scoring along with them, and then giving interviews that were scored by their companion as well until over 90% agreement was reached for three consecutive interviews carried out by the new interviewer. During data collection for the validity study, training tapes for the ADI-R became available. All interviewers coded these tapes, blind to diagnosis, with percentage agreement for exact score completed item by item consistently above 90%, and averaging about 94%. Over the course of the study, reliability checks were made by having two research associates conduct approximately one in five interviews together.

Parents of all but 5 of the additional autistic children and 7 of the additional mentally handicapped/language-impaired children were interviewed as part of an outpatient diagnostic assessment. Interviewers were unfamiliar with the child but did have some background information including reasons for referral and previous diagnoses. Since typically about

40% of children referred to the clinics known for seeing autistic children were diagnosed as not autistic, and about 10% of the children referred to the clinic for developmental disorders were diagnosed as autistic, this was felt to be similar to interviewers remaining blind. Only children who later received DSM-III diagnoses of either autism, language impairment (LI), or mental handicap (MH), without autism/PDD made by a psychiatrist and/or psychologist who was unfamiliar with the ADI-R score were included in the study. For the validity study, consensus scores for the reliability subjects were created through discussion among the interviewers and raters.

Algorithm

The appropriateness of the algorithm generated for the ADI-R was tested with the preschool sample by comparing individual items and area summary scores across diagnostic groups and by comparing the number of children who met formal criteria for autism. One-way fixed-effect ANOVAs (2 diagnoses) were performed on all item scores and algorithm area scores. Because variances were unequal, scores were converted to ranks before analyses were performed. Raw scores are reported for all measures for the sake of interpretability. Because of the large number of analyses, only those with values of $p < .02$ are treated as significant. When computing the diagnostic algorithm, items coded 3 for particularly severe manifestations or 7 for indications of abnormality that differed from the dimension in question were recoded as 2 and 0, respectively. Only items appropriate for children 4 years and under were analyzed in the validity studies because of the young age of the sample.

Social

As shown in Table II, all 13 algorithm items from the social area that could be scored in preschool children showed significant diagnostic differences across autistic and mentally handicapped groups. Several nonalgorithm items that described behaviors typically attributed to autistic children were not significant, such as *cuddliness* and *discrimination of parents.* Though the diagnostic group difference for current *separation anxiety* was significant; nearly one third of the autistic children were scored as not abnormal (8/25 did show separation anxiety) or mildly abnormal (11 out of 25).

Table II. Mean Scores for ADI-R Algorithm Items for Reciprocal Social Interaction

	Autistic (n = 25)		Mentally handicapped/ Language-impaired (n = 25)		
	M	*SD*	*M*	*SD*	ANOVAs: $F(1, 48)$
B1. Failure to use eye-to-eye gaze, facial expression, body posture and gesture to regulate social interaction					
Direct gaze	0.92	0.76	0.36	0.64	7.97[a]
Social smiling	1.28	0.89	0.20	0.50	27.95[b]
Range of facial expression used to communicate	1.32	0.69	0.20	0.41	48.75[b]
B2. Failure to develop peer relationships					
Interest in children	1.72	0.54	0.44	0.58	64.67[b]
Response to children	1.40	0.64	0.36	0.57	36.54[b]
B3. Lack of seeking to share own enjoyment					
Showing and directing attention	1.84	0.47	0.40	0.71	71.67[b]
Offering to share	1.68	0.56	0.76	0.88	19.53[b]
Seeking to share own enjoyment	1.48	0.65	0.32	0.63	41.02[b]
B4. Lack of social emotional reciprocity and modulation to context					
Use of other's body	1.44	0.82	0.20	0.41	45.76[b]
Inappropriate facial expressions	1.16	0.80	0.12	0.33	36.05[b]
Quality of social overtures	1.64	0.49	0.40	0.58	67.05[b]
Appropriateness of social responses	1.64	0.64	0.40	0.58	51.95[b]
Offers comfort	1.48	0.71	0.04	0.20	94.25[b]

[a] $p < .01$.
[b] $p < .001$.

As shown in Table III, all social subdomain totals, based on ICD-10/DSM-IV draft guidelines (e.g., integration of nonverbal behaviors, peer relationships), also yielded significant differences between the autistic and MH/LI groups. In a one-way fixed-effect ANOVA performed on the total social algorithm scores converted to ranks, autistic subjects had significantly higher scores (indicating greater abnormality) than the MH/LI children, $F(1, 48) = 243.38$, $p < .0001$.

For the social algorithm total, 2 autistic children had a score of 11 (10 was the cutoff); the remaining 24 autistic children had scores of 13 or greater. In contrast, no children from the MH/LI group had social domain scores exceeding 10 out of 29. Two MH/LI children had scores of 10.

Communication

Two subdomains of communication were coded for children at all language levels. These were the areas of gesture (C1) and play (C4). These subdomains are discussed first and then the remaining subdomains considered separately for children with phrases (verbal) and children with single words only or no words (nonverbal).

As shown in Table IV, there were significant main effects for diagnosis for all gesture (C1) items. Autistic youngsters were described as using fewer of all types of gestures than the MH/LI children. The three items in the subdomain of play (C4), *spontaneous imitation, imaginative play,* and *imitative social play,* all yielded significant main effects of diagnosis. Children with autism showed higher (more abnormal) scores in all aspects of play and imitation.

Items from the two subdomains of verbal communication were compared across diagnosis for children with phrases only. These domains were conversation (CV2), and stereotyped use of language (CV3). Out of six items, there were significant main effects of diagnosis for only one: *social chat.* For the five other items, scores for the autistic children indicated greater abnormalities; however, because of small samples and floor effects on some items in this relatively young sample, differences were not significant.

Several nonalgorithm communication items also revealed diagnostic group differences. The autistic children acquired their first words at a later age and were described by their parents as less likely to babble in a socially directed way as infants than the MH/LI children. They also differed in their use of enactive and emotional gestures, comprehension of language, understanding gestures, and elicited vocal imitation. Significant diagnostic differences for communication items administered only for the verbal children occurred in immediate echolalia and understanding the plots of simple stories.

Table III. Means and Standard Deviations for ADI-R Subdomains and Domain Scores

	Autistic (n = 25)		Mentally handicapped/ Language-impaired (n = 25)		
	M	*SD*	*M*	*SD*	ANOVAs
Quantitative Impairments in Reciprocal Social Interaction					
B1. Failure to use eye-to-eye gaze, facial expression, body posture, and gesture to regulate social interaction	3.52	1.61	0.76	1.01	$F(1, 48) = 52.66^c$
B2. Failure to develop peer relationships	3.12	0.93	0.80	1.04	$F(1, 48) = 69.24^c$
B3. Lack of seeking to share own enjoyment	5.00	1.26	1.48	1.58	$F(1, 48) = 75.67^c$
B4. Lack of social emotional reciprocity and modulation to context	7.36	1.93	1.16	1.21	$F(1, 48) = 184.34^c$
B. Social domain total	19.00	3.76	4.20	2.88	$F(1, 48) = 243.38^c$
Quantitative Impairments in Communication and Language					
C1. Delay or total lack of spoken language not compensated by gesture	6.60	1.47	2.08	2.50	$F(1, 48) = 60.73^c$
C2V. Relative failure to initiate or sustain conversational interchange[a]	3.22	1.09	1.84	1.52	$F(1, 21) = 19.40^c$
C3V. Stereotyped and repetitive use of language[a]	3.11	1.90	1.36	0.63	$F(1, 21) = 10.38^a$
C4. Lack of varied spontaneous make-believe or social imitative play	3.44	1.51	1.84	1.52	$F(1, 48) = 38.44^c$
CNV. Communication total (nonverbal children)[a]	11.62	1.96	5.09	4.28	$F(1, 25) = 28.91^c$
CV. Communication total (verbal children)[a]	16.33	2.96	5.57	3.06	$F(1, 21) = 69.60^c$

Table IV. Mean Scores for ADI-R Algorithm Items for Communication[a]

	Autistic (n = 25)		Mentally handicapped/ Language-impaired (n = 25)		
	M	*SD*	*M*	*SD*	ANOVAs
C1. Delay or total lack of spoken language, not compensated by gesture					
Pointing to express interest	1.68	0.47	0.48	0.87	$F(1, 48) = 36.49^{c}$
Conventional gestures	1.72	0.46	0.60	0.87	$F(1, 48) = 32.67^{c}$
Nodding head	1.80	0.50	0.68	0.85	$F(1, 48) = 32.11^{c}$
Head shaking	1.40	0.76	0.32	0.56	$F(1, 48) = 32.64^{c}$
C2. Relative failure to initiate or sustain conversational interchange[a]					
Social chat	1.55	0.53	0.28	0.47	$F(1, 21) = 36.52^{c}$
Reciprocal conversation	1.67	0.71	0.93	0.83	ns
C3V. Stereotyped and repetitive use of language					
Stereotyped utterances	1.55	0.73	0.71	0.73	ns
Inappropriate questions	0.44	0.73	0.14	0.36	ns
Pronominal reversal	0.87	0.76	0.07	0.25	ns
Neologisms/idiosyncratic language	0.22	0.44	0.07	0.27	ns
C4. Lack of varied spontaneous make-believe or social imitative play					
Spontaneous imitation	1.72	0.54	0.68	0.85	$F(1, 48) = 26.51^{c}$
Imaginative play	1.48	0.77	0.68	0.80	$F(1, 48) = 12.94^{b}$
Imitative social play	1.24	0.66	0.48	0.51	$F(1, 48) = 20.63^{c}$

[a] n = 16 For nonverbal autistic and n = 11 for nonverbal MH/LI; n = 9 for verbal autistic and n = 14 for verbal MH/LI.
[b] $p < .01$.
[c] $p < .001$.

As shown in Table III, algorithm summary scores for communication for verbal children yielded a significant main effect for diagnosis, $F(1, 21) = 69.6, p < .0001$. All 9 verbal autistic youngsters fell in the range of autism (10 or greater); 5/14 verbal MH/LI children also scored in this range. There was also a significant effect of diagnosis for the nonverbal children, $F(1, 25) = 28.91, p < .001$. All nonverbal autistic children met the communication criterion, as did 4 MH/LI children.

Restricted and Repetitive Behaviors

As shown in Table V, four items from the ICD-10/DSM-IV domain of restricted and repetitive interests and behaviors showed significant differences in distribution for autistic and nonautistic children. These included *verbal rituals, unusual sensory behaviors, hand and finger mannerisms,* and *whole body mannerisms. Repetitive use of objects* did not show a significant difference; *compulsions/rituals* and *unusual preoccupations* were marginally significant. However, because of findings from previous research with older populations of the diagnostic usefulness of these items (Le Couteur et al., 1989; Lord et al., in press), they were retained in the algorithm. They require continued evaluation, however. Nonalgorithm items such as *abnormal idiosyncratic negative response, unusual attachments to objects, resistance to change in own routine,* and response to trivial changes in the environment did not differ across diagnosis.

The mean score for the algorithm area of restricted, repetitive behaviors was 4.92 (SD = 1.80) for the autistic children, compared to 1.96 ($SD = 1.64$) for the mentally handicapped children/language impaired children, $F(1, 48) = 37.56, p < .0001$ (performed on ranks). Of the 25 autistic children, 1 had a score of 2 (not meeting criteria), 7 children scored a total of 3 in this area; the remaining 17 autistic children scored 4 or more. Of the 25 MH/LI children, 6 received scores of 3 or higher.

Items Outside Diagnostic Areas

Few differences were significant for nonalgorithm items outside the three diagnostic areas. Only *curiosity, initiation of appropriate activities, food fads, sensitivity to noise, self-injurious behavior,* and *unusual fears* were reliably different across diagnostic categories. Severity of aggressive behavior, pica, tantrums, motor difficulties, and sleep problems did not distinguish the diagnostic groups.

Table V. Mean Scores for ADI-R Algorithm Items for the Area of Restricted, Repetitive Behavior and Interests

	Autistic (n = 25)		Mentally handicapped/ Language-impaired (n = 25)		
	M	*SD*	*M*	*SD*	ANOVAs: $F(1, 48)$
D1. Encompassing preoccupations					
Unusual preoccupations	1.04	0.93	0.48	0.77	5.35[a]
D2. Apparently compulsive adherence to nonfunctional rituals					
Verbal rituals	1.16	1.21	0.40	0.87	14.26[b]
Compulsions/rituals	0.68	0.80	0.20	0.50	6.45[a]
D3. Stereotyped and repetitive motor mannerisms					
Hand and finger mannerisms	1.20	0.76	0.20	0.58	27.27[c]
Other complex mannerisms	0.88	0.93	0.24	0.60	8.42[b]
D4. Preoccupation with part-objects or nonfunctional elements of materials					
Repetitive use of objects	1.00	0.76	0.72	0.68	ns
Unusual sensory interests	1.12	0.60	0.36	0.57	21.13[b]

[a] $p < .02$.
[b] $p < .01$.
[c] $p < .001$.

Restricted, Repetitive Behaviors and Interests					
D1. Encompassing preoccupations	1.04	0.93	0.48	0.77	ns
D2. Apparently compulsive adherence to nonfunctional rituals	1.16	1.21	0.40	0.87	ns
D3. Stereotyped and repetitive motor mannerisms	1.36	0.76	0.32	0.69	$F(1, 48) = 25.75^{c}$
D4. Preoccupation with part-objects and nonfunctional elements of materials	1.36	0.57	0.76	0.66	$F(1, 48) = 11.79^{b}$
D. Restricted repetitive behaviors and interests total	4.92	1.80	1.96	1.64	$F(1, 48) = 37.56^{c}$

[a] $n = 16$ For nonverbal autistic and $n = 11$ for nonverbal MH/LI; $n = 9$ for verbal autistic and $n = 14$ for verbal MH/LI.
[b] $p < .01$.
[c] $p < .001$.

Overall Algorithm

All but one of 25 clinically diagnosed autistic children met ADI-R criteria for a ICD-10/DSM-IV draft diagnosis of autism. The one exception was a high-functioning 4-year-old boy with autism who received a score of 2 for *restricted, repetitive behaviors* (one less that the cutoff). Two for the 25 mentally handicapped/language impaired children were classified as autistic by the ADI-R, but did not receive clinical diagnoses of autism. Both met criteria exactly or by 1 point in two out of three areas; both boys were nonverbal, young 3-year-olds with mental ages of 21 and 26 months, respectively. They had received clinical diagnoses of general developmental delay, severe receptive-expressive language impairment, and in one case, attention deficit/hyperactivity disorder and in the other, oppositional disorder.

DISCUSSION

The ADI-R is a reliable and valid instrument for making diagnoses of autistic children of preschool age. Interrater reliability is good, with kappas ranging from .62 to .89 and equivalent to those found for the original ADI for communication and social items. Interrater reliability for items in the area of restricted and repetitive behaviors and interests is adequate, with a mean kappa of .70, but lower than that found with the earlier version of the ADI, in part because of slightly lower kappas for *compulsions and rituals, unusual preoccupations, hand and finger mannerisms,* and both resistance to change items. Percentage agreement for each of these items was over 90%, but frequencies of occurrence were sufficiently low to restrict the range of kappas. Since the present data were collected by interviewers less experienced with autism (though given substantial experience with the ADI-R) from more naive parents of children with a mean age of $3\frac{1}{2}$-4 years, these findings are encouraging, particularly for clinical use of the instrument. Further samples of older children and adults require study as well.

As for the earlier version of the ADI, intraclass correlations are very high. Internal consistency is also quite good, particularly for communication and social items. Reliability across time is adequate but needs to be studied further with a larger sample.

Individual items were much better at discriminating autistic from mentally handicapped/language-impaired preschool children than had been expected. Given earlier findings that "ever" scores and scores that targeted behaviors during the 4- to 5-year period were most discriminative, the finding

that all social and nonverbal communication algorithm items yielded significant differences for diagnosis was surprising. The algorithm items that did not differentiate young children included some verbal communication items, for which results were limited by the small samples, and several items from the area of restricted and repetitive behaviors that may have been affected by the young age of the subjects. When clear differences had been found for the item in the DSM-IV field trials (Lord et al., in press) or for the earlier version of the item with the older sample and the item had been changed very little from the ADI, the decision was made to keep it in the algorithm.

Conclusions

This paper describes a revision of the Autism Diagnostic Interview, the Autism Diagnostic Interview–Revised (ADI-R), a semistructured, investigator-based interview for caregivers of children and adults for whom autism or pervasive developmental disorders is a possible diagnosis. The revised interview has been reorganized, shortened, ad modified to be appropriate for children with mental ages from about 18 months into adulthood. It is now linked more closely to ICD-10 and DSM-IV criteria. Its psychometric properties are strong for preschool children but require additional investigation, particularly with older children and adults. Further comparisons between well-defined autistic samples and other groups with pervasive developmental disorders and related difficulties will also be important in indicating the usefulness of the ADI-R in diagnosing autism versus other pervasive developmental disorders.[3]

[3]Use of the ADI-R for research purposes requires training in both administration and scoring by a person experienced in use of the instrument who has established reliability with other experienced individuals. Training workshops are now available annually in the U.K. and in North America. Training videotapes that describe the organization and purpose of the ADI-R and provide detailed examples of administration and scoring of social, communication, and restricted and repetitive interest items are now available as part of the training package. These tapes allow trainees to compare their codings with those of the consensus codes of the authors. Each tape is provided with a commentary and justifications for decisions. Before using the ADI-R in a research project, investigators are requested to attend a training workshop, to show reliability with consensus codings for at least two standard interviews and to demonstrate the ability of another trained researcher to reliably score two examples of interviews administered by the investigator.

APPENDIX
Nonalgorithm items for ADI-R: Reliability and Validity with a Preschool Sample[a]

	κ_w	Aggrement	F(1, 48)
Social and Play Items			
Arms up to be lifted (infant)	.52	.90	11.82[c]
Arms up to be lifted (current)	.80	.90	7.40[b]
Infant direct gaze	.89	.95	ns
Infant separation anxiety	.78	.90	9.73[b]
Separation anxiety (current)	.62	.90	8.79[b]
Cuddly as infant	.76	.91	ns
Cuddly as child (current)	.75	.90	ns
Infant social smile	.72	.89	ns
Social smile at 2 years	.75	.89	18.42[c]
Secure base	.81	.97	5.69[b]
Attention to voice (current)	.59	.87	14.66[c]
Affection	.85	.94	9.09[b]
Come for comfort at 2 years	.80	.93	15.77[c]
Come for comfort (current)	.68	.91	11.68[c]
Discrimination of parents	.54	.96	ns
Sense of humor	.78	.90	11.69[c]
Join others' activities	.72	.90	15.57
Sharing others' pleasure	.85	.95	21.43[c]
Greeting	.81	.92	18.61[c]
Social disinhibition	.70	.94	ns
Communication Items All Levels			
Form of babble as an infant	.70	.91	5.80[b]
Age in months at first words	.89	.94	7.72[b]
Age in months at first phrase	.89	.95	ns
Emotional gestures	.74	.90	17.90[c]
Enactive gestures	.71	.93	9.37[b]
Elicited vocal imitation	.69	.90	9.69[b]
Reciprocal vocalization	.77	.93	ns
Understanding gesture	.82	.93	23.35[c]
Comprehension of language	.79	.92	33.94[c]
Verbal Only			
Amount of social language	.89	.94	ns
Immediate echolalia	.85	.95	5.32[b]
Gestures accompanying speech	.77	.91	ns

Report of events	.81	.92	ns
Talk expressing interest in others	.72	.87	ns
Vocal expression	.81	.91	ns
Understanding plot	.76	.92	1.82[c]
Intonation, rhythm, rate	.72	.92	ns
Other			
Unusual attachments to objects	.64	.90	ns
Aggression to family member	.75	.89	ns
Aggression to non-family members	.77	.93	ns
Tantrums	.96	.98	ns
Destructive behaviors	.83	.94	ns
Self-injury	.96	.99	5.96[b]
Food fads	.75	.86	5.77[b]
Pica	.94	.97	ns
Unusual fears	.65	.78	8.16[b]
Cry because of pain	.65	.81	ns
Cry for social reasons	.68	.81	ns
Faint or seizures	.72	.95	ns
Fine motor skills	.83	.96	ns
Gait	.78	.92	ns
Clumsiness	.57	.42	ns
Initiation of appropriate activities	.85	.95	9.94[b]
Curiosity	.82	.97	10.63[b]
Unusual musical ability	.64	.56	ns
Abnormal idiosyncratic negative response	.30	.93	ns
Distress over changes in own routine	.77	.91	ns
Distress over trivial changes in enviornment	.86	.96	ns
Unusually good memory	.83	.93	ns
Overactivity at home	.73	.90	ns
Problems going to bed	.77	.93	ns
Sleep problems	.79	.92	ns
Rocking	.90	.97	ns
Sensitive to noise	.83	.91	5.62[b]
Overall level of language	.92	.97	ns
Concerns about hearing	.66	.84	ns

[a] All items refer to current behavior (in last 3 months) unless otherwise noted. Reliability analyses (k_w, agreement) were performed on 11 subjects for verbal items and 20 for all others. Validity analyses (F scores) were performed using 30 subjects for verbal items and 50 for all others. F scores are for main effect of diagnosis performed on ranks.

[b] $p < .01$.

[c] $p < .001$.

REFERENCES

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

American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

American Psychiatric Association. (1993). *Options book for DSM-IV.* Washington, DC: Author.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind?" *Cognition, 21,* 37-46.

Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. (Eds.). (1993). *Understanding other minds: Perspectives on the theory of mind hypothesis of autism.* Oxford, U.K.: Oxford University Press.

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

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement of partial credit. *Psychological Bulletin, 70,* 213-220.

Conger, A. J. (1980). Integration and generalization of kappas for multiple raters.

Frith, U. (1989). *Autism: Explaining the enigma.* Oxford, U.K.: Basil Blackwell.

Hagberg, M. M., Aicardi, J., Dias, K., & Rasmos, O. (1983). A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett syndrome: Report of 35 cases. *Annals of Neurology, 14,* 471-479.

Le Couteur, A., Rutter, M., Lord, C., Rios, P., Robertson, S., Holdgrafer, M., & McLennan, J. D. (1989). Autism Diagnostic Interview: A semistructured interview for parents and caregivers of autistic persons. *Journal of Autism and Developmental Disorders, 19,* 363-387.

Lord, C., Pickles, A., McLennan, J., Rutter, M., Bregman, J., Folstein, S., Fombonne, E., Leboyer, M. and Minshew, N. (in press). Diagnosing autism: Analyses of data from the Autism Diagnostic Interview. In A. Frances (Ed.), *DSM-IV sourcebook.* Washington: APA.

Lord, C., & Rutter, M. (1994). Autism and pervasive developmental disorder. In M. Rutter, E. Taylor, & L. Hersov (Eds.), *Child and adolescent psychiatry* (3rd ed.) (pp. 569-593). Oxford: Blackwell Scientific.

Lord, C., Storoschuk, S., Rutter, M., and Pickles, A. (1993). Using the ADI-R to diagnose autism in preschool children. *Infant Mental Health Journal, 14,* 234-252.

Olsson, B., & Rett, A. (1990). A review of Rett syndrome with a theory of autism. *Brain and Development, 12,* 11-15.

Short, C. B., & Schopler, E. (1988). Factors relating to age of onset in autism. *Journal of Autism and Developmental Disorders, 18,* 207-216.

Stutsman, R. (1931). Guide for administering the Merrill-Palmer Scale of Mental Tests. In L. M. Terman (Ed.), *Mental measurement of preschool children* (pp. 139-262). New York: Harcourt, Brace & World.

Treiman, D. J. (1977). *Occupational prestige in comparative perspectives.* New York: Academic Press.

Tsai, L. Y. (1992). Is Rett syndrome a subtype of pervasive developmental disorders? *Journal of Autism and Developmental Disorders, 22,* 551-561.

Volkmar, F. R., Klin, A., Siegal, B., Szatmari, P., Lord, C., Cicchetti, D. V., & Rutter, M. (in press). DSM-IV autism/pervasive development disorder field trial final report. *American Journal of Psychiatry.*

Wolff, P., Gardner, J., Paccia, J., & Lappen, J. (1989). The greeting behavior of Fragile X males. *American Journal on Mental Retardation, 93,* 406-411.

World Health Organization. (1987). *ICD-10 1986 draft of chapter 5 categories F00-F99. Mental, behavioral and developmental disorders.* Geneva: Author.

World Health Organization. (1992). *The ICD-10 classification of mental and behavioral disorders: Clinical descriptions and diagnostic guidelines.* Geneva: Author.