J Autism Dev Disord (2007) 37:49–61
DOI 10.1007/s10803-006-0334-4

ORIGINAL PAPER

The First Year Inventory: Retrospective Parent Responses to a Questionnaire Designed to Identify One-Year-Olds at Risk for Autism

Linda R. Watson · Grace T. Baranek · Elizabeth R. Crais · J. Steven Reznick · Jessica Dykstra · Twyla Perryman

Published online: 12 January 2007

Abstract The First Year Inventory (FYI) is a parent questionnaire designed to assess behaviors in 12-month-olds that suggest risk for an eventual diagnosis of autism. We examined the construct validity of the FYI by comparing retrospective responses of parents of preschool children with autism spectrum disorders (ASD; $n = 38$), other developmental disabilities (DD; $n = 15$), and typical development (TD; $n = 40$). Children with ASD were rated at significantly higher risk on the FYI than children with DD or TD. The DD group was at intermediate risk, also significantly higher than the TD group. These retrospective data strengthen the validity of the FYI and have implications for refining the FYI to improve its utility for prospective screening of 12-month-olds.

Keywords Autism · Infants · Screening · Parent questionnaire · Social-communicative · Sensory-regulatory

L. R. Watson (✉) · E. R. Crais · J. Dykstra · T. Perryman
Division of Speech and Hearing Sciences, School of Medicine, The University of North Carolina, CB# 7190, Chapel Hill, NC 27599-7190, USA
e-mail: Linda_Watson@med.unc.edu

G. T. Baranek
Division of Occupational Sciences, School of Medicine, The University of North Carolina, Chapel Hill, NC, USA

J. Steven Reznick
Psychology Department, The University of North Carolina, Chapel Hill, NC, USA

Present Address:
J. Dykstra
Kennedy Krieger Institute, Baltimore, MD, USA

Introduction

Efforts to develop and validate broad-based screening tools to identify very young children at risk for autism spectrum disorders (ASD) have been ongoing since the pioneering efforts of Baron-Cohen and colleagues (Baron-Cohen, Allen & Gillberg, 1992). Successful screening would create opportunities for prospective research on bio-behavioral processes involved in ASD and investigations of the efficacy of early intervention. From a public health perspective, early intervention for children at risk for ASD requires valid and cost-effective identification of children who could benefit from these early services.

The initial efforts at broad-based screening took place in the United Kingdom. The Checklist for Autism in Toddlers (CHAT; Baron-Cohen et al., 1992, 1996) was designed to identify 18-month-olds at risk for autism, and included nine *yes/no* questions for parents and five items that could be observed by home health visitors. Baird et al. (2000) conducted a six-year follow-up on the implementation of the CHAT for general population screening and found high specificity (most children who passed the CHAT did not receive a subsequent diagnosis of autism) and positive predictive value (most children who failed the CHAT were subsequently diagnosed with autism), but low sensitivity (many children who passed the CHAT eventually developed autism).

Robins, Fein, Barton, and Greene (2001) adapted the CHAT to eliminate the observer-report questions and expand the number of parent-report questions to 23, thus giving the instrument more utility in general population screenings in the United States. The Modified CHAT (MCHAT) focused on 24-month-olds

rather than on 18-month-olds to improve its sensitivity as well as its acceptability to health care providers. To evaluate the MCHAT, Robins et al. collected data on an unselected pediatric sample of 1122 18- to 25-month-old children and 171 high-risk 18- to 30-month-old children referred to the early intervention system in Connecticut. Children who passed the MCHAT received no follow-up. Based on phone calls to the parents of 132 children who failed the MCHAT, the researchers determined that 58 children should be further evaluated. Nineteen of the 58 children had developmental delays but not an ASD, and 39 were diagnosed with an ASD. Validation of the MCHAT, including large scale screening and follow-up of an unselected population, is continuing (Robins & Dumont-Mathieu, 2006).

Another tool designed for general population screening of children from 12 to 48 months is Stage 1 of the Pervasive Developmental Disorder Screening Test-II (PDDST-II) (Siegel, 2004). Siegel reports that the PDDST-II has high sensitivity for identifying children in high risk samples in need of further evaluation, but no data have been reported from an unselected population.

Finally, two recent reports describe the development and validation of the Early Screening for Autistic Traits questionnaire (ESAT; Dietz, Swinkels, van Daalen, van Engeland, & Buitelaar, 2006; Swinkels et al., 2006). The ESAT is a 14-item caregiver questionnaire developed for use with 14–15 month olds. Initial findings suggested that four items would be useful for prescreening (Swinkels et al.). These four prescreening items were used in a population screening of more than 31,000 infants attending well-baby clinics in the province of Utrecht, the Netherlands (Dietz et al.). The researchers reported on results up to the age of 42 months for infants who failed the prescreening. Through the ESAT, 5.7 children per 10,000 were identified with ASD. The authors did not include follow-up information on children who passed the prescreening, thus precluding the calculation of sensitivity and specificity for the ESAT; however, they note that a comparison of their data to recent prevalence figures suggests that the ESAT missed many children at risk for ASD. For example, recent prevalence estimates from the Centers for Disease Control and Prevention (2006) indicate that approximately 60 children per 10,000 will eventually be diagnosed with ASD. The ESAT also yielded a large number of false positives (i.e., 57 of the 73 children available for follow-up diagnostic evaluations). Notably, all false positives eventually resulted in a diagnosis of a developmental disability, but not ASD.

Efforts to screen infants for risk of ASD have raised a number of issues. One issue is determining the earliest age at which identification of autism risk may be feasible, and the optimal age for screening to maximize the sensitivity, specificity, and positive and negative predictive value of the screening process. A second issue is whether the format of the screening process allows effective implementation consistent with goals and resources of public health policy. A final issue is the impact of poor sensitivity and false negatives, which would leave many at-risk infants undiagnosed and untreated, and poor specificity and false positives, which would induce unnecessary stress and anxiety.

The efforts of our team to develop an early screening tool have been based on the premise that screening for autism risk is feasible as early as 12 months of age. We cannot yet claim that this age is optimal, but we accept the principle that early detection provides considerable advantage. A growing corpus of research indicates that differences between infants with ASD and those with other developmental disabilities (DD) as well as typical development (TD) are apparent at least by 9–12 months of age (e.g., Baranek, 1999; Maestro, Casella, Milone, Muratori, & Palacio-Espasa, 1999; Osterling, Dawson, & Munson, 2002; Zwaigenbaum et al., 2005). Significant group differences, however, do not necessarily mean that the diagnostic outcomes for individual children can be predicted reliably based on infant behaviors. Even when the measures made in infant research studies predict later diagnostic outcomes, those measures are not always easily translated for widespread, efficient clinical use. Nevertheless, studies of infants with ASD by methods including retrospective video analysis (e.g., Baranek, 1999, Maestro et al., 1999, Osterling et al., 2002) and prospective studies of infant siblings of children with ASD (e.g., Landa & Garrett-Mayer, 2006; Yirmiya et al., 2006; Zwaigenbaum et al., 2005) have provided valuable information on symptoms and development in infants later diagnosed with ASD.

Of particular relevance to the current study, Bryson, McDermott, Rombough, Brian, and Zwaigenbaum (Unpublished) developed the Autism Observation Scale for Infants (AOSI) for use with 6- to 12-month olds. During administration of the AOSI, examiners look for 18 potential risk markers, including visual attention and attention disengagement, coordination of eye gaze with action, imitation, affect, behavioral reactivity, social-communicative behaviors, and sensorimotor development. In a study of infant siblings of children with autism, Zwaigenbaum et al. (2005) found that the diagnostic status of those siblings at 24 months

was predicted by the number of risk markers they exhibited on the AOSI at 12 months. This study supports the assumption that risk for autism can be identified, at least in some children, at 12 months. Unfortunately, an observational assessment of this type is labor-intensive and costly, decreasing its utility for large-scale screenings in unselected populations of infants.

To translate the information on the characteristics of infants later diagnosed with ASD into a tool that would be appropriate for large-scale screening, our research team developed a parent questionnaire, the First Year Inventory (FYI; Baranek, Watson, Crais, &, Reznick, Unpublished manuscript; Reznick, Baranek, Reavis, Watson, & Crais, in press). The project began with an extensive review of the literature on infants and toddlers with ASD. We generated a list of characteristics reported to distinguish infants and toddlers eventually diagnosed with ASD, focusing on characteristics that would be identifiable in 12-month-old infants. These characteristics included unusual behaviors, the absence of typical behaviors, and typical behaviors occurring only with extensive parental support.

Parent-report questions were developed to assess each identified characteristic. The FYI was revised multiple times based on feedback from parents of children with ASD and professionals, and three small-scale mailings to parents of 12-month-olds in a central region of North Carolina (NC). The FYI Version 2.0 consists of a total of 63 questions: 46 with response alternatives ''never'', ''seldom'', ''sometimes'', and ''often''; 14 with three or four ad hoc multiple choice answers; one in which the parent indicates sounds produced by the infant from a list of consonants; and two open-ended questions about parental concerns and unusual physical or medical characteristics. Normative data on the FYI Version 2.0 were obtained from a mailing of 5941 FYIs, resulting in 1496 responses from parents of 12-month olds (Reznick et al., in press). These data were used to derive an initial scoring of the FYI, based on assigning risk points for answers with an extremely low frequency. Using exploratory factor analysis and construct shaping procedures, we derived 8 separate constructs. Four constructs are in the Social-Communication domain: Social Orienting and Receptive Communication, Social Affective Engagement, Imitation, and Expressive Communication. Four additional constructs are in the Sensory-Regulatory domain: Sensory Processing, Regulatory Patterns, Reactivity, and Repetitive Behavior.

One dilemma at the early stage of validating a tool for the early identification of ASD is finding a sample with sufficient numbers of infants with ASD to yield meaningful evaluation of the tool's discriminative performance. For the current study, our strategy was to recruit parents of older children already diagnosed with autism or other developmental disabilities as well as children with typical development. Then we asked parents to complete a retrospective version of the FYI to determine how well the FYI could discriminate among these known diagnostic groups, and to acquire information that will help us refine the instrument. The disadvantage of this strategy is that it prohibits calculating true rates of sensitivity and specificity. The advantage is that face validity can be assessed, a necessary first step toward a full-scale assessment of the tool's clinical utility. We hypothesized that our three groups would differ significantly on overall risk score, individual item risk scores, and risk scores on the eight constructs as well as the two construct domains (Social-Communicative and Sensory-Regulatory), with the autism group showing the highest level of risk. We also examined the statistical properties of the FYI (sensitivity, specificity, positive predictive value, and negative predictive value) as applied to our retrospective sample. Within the ASD group, we explored the relationship between risk scores and age of diagnosis, age of first concerns, reported regression, and preschool symptom severity.

Methods

Participants

We recruited three groups of preschoolers: 38 with ASD, 15 with a DD not in the autism spectrum, and 40 exhibiting typical development (TD). An additional 15 children were excluded from the study because they did not meet criteria for any of the three groups, as detailed below. Descriptive data for the participants in these three groups are provided in Table 1.

Recruitment

We recruited participants in various ways. Most participants with ASD were recruited through the Autism Research Subject Registry of the Neurodevelopmental Disorders Research Center at The University of North Carolina at Chapel Hill (UNC-CH). Many participants in all three groups were recruited through a collaborating research study. We also recruited participants through mailings to NC chapters of The Arc of the United States, diagnostic and therapeutic clinics in NC, and area child care programs. Finally, we recruited a small number of participants through word-of-mouth.

Table 1 Description of participants

	TD	DD	ASD
n	40	15	38
Mean age in months (SD)	39.3 (12.0)	45.1 (18.9)	46.1 (15.9)
Range	20–64	20–75	14–73
Mullen Standard Score (SD)	110.0 (n = 10)	62.6 (n = 12)	60.2 (n = 19)
Range	96–126	49–85	49–99
% male	52.5	86.7	86.8
% first born	51.3	54.1	28.6
% preterm (<37 weeks)	13.5	13.3	20.6
% birth complications	7.5	60.0	30.6
% family history of any DD	12.5	26.7	44.1
% white, non-Latino	92.5	100.0	84.4
% mothers with =/> bachelor's degree	95.0	66.7	57.5

Interested parents contacted us by phone, response card, or email.

Parents expressing interest received a packet containing an explanatory letter, consent form, and the FYI-Retrospective. Parents participating by mail also received other parent questionnaire assessment instruments, as described below. Parents returned the forms either directly to the researchers or by using a postage-paid pre-addressed envelope.

Instruments

The *First Year Inventory-Retrospective Version (FYI-R)*

All parents in the study completed the *FYI-Retrospective Version* (FYI-R), which is the FYI (Version 2.0) rewritten into the past tense. For example, the FYI question ''Does your baby seem to have trouble hearing?'' was replaced with ''Did your baby seem to have trouble hearing?'' The opening explanation of the FYI-R included prompts to help parents focus on what their child was like at his or her first birthday. One FYI question, which asks about specific sounds that the child was making at 12 months of age, was eliminated because it seemed unlikely that parents could give reliable retrospective reports on that item. In addition, the two open-ended questions asking about developmental concerns were omitted because children in the two clinical groups already had known disabilities.

For participants recruited through the collaborating research study, we had extensive information collected from direct observation and assessment of the child as well as caregiver questionnaires and interviews with caregivers. For the purposes of the present study, we used data from the following instruments:

Demographic Data and Medical History Interviews: Research staff collected demographic data and medical history from families by interview. Demographics included parent education, employment, income, race and ethnicity, and family composition. Medical history included birth history, parent concerns about the target child, age(s) of any diagnoses, source(s) of diagnoses, known genetic or other medical conditions, medications, and family history of DD/ASD.

The *Mullen Scales of Early Learning* (MSEL) (Mullen, 1995) is a standardized, comprehensive measure of development for infants and preschool children from birth to 68 months, consisting of five subscales: visual reception, gross and fine motor, receptive and expressive language.

The *Leiter International Performance Scale-Revised* (Leiter-R) (Roid & Miller, 1997) is a standardized nonverbal measure of intelligence. We used the ''Brief IQ'', based on four subtests (Repeated Patterns, Sequential Order, Figure-Ground, and Form Completion). The Leiter-R served as a cognitive measure for children too old for assessment with the MSEL.

The *Vineland Adaptive Behavior Scale—Survey Edition* (VABS) (Sparrow, Balla & Cicchetti, 1984) is a standardized caregiver interview instrument designed to evaluate children aged 0–18 years for current adaptive behavior in four areas (communication, daily living, socialization, motor).

The *Autism Diagnostic Interview—Revised* (ADI-R) (Rutter, LeCouteur, & Lord, 2003) is a standardized parent interview designed to assess the presence and severity of symptoms based on the diagnostic criteria for autism in the Diagnostic and Statistical Manual of Mental Disorders (4th edition; DSM-IV—American Psychiatric Association, 1994). The ADI-R assesses three categories of symptoms: deficits in social relatedness, deficits in communication, and restricted range of interests and behaviors. In addition, the ADI-R includes questions about age of initial symptoms and developmental regression.

The *Autism Diagnostic Observation Schedules* (ADOS) (Lord, Rutter, DiLavore & Risi, 1999) is a standardized 30–45 minute observational assessment. Research staff administered the ADOS to children in the ASD and DD groups. The ADOS consists of four modules, each for a different level of language ability. Modules 1 (designed for children who are preverbal or

just beginning to speak) and 2 (designed for children at a phrase or simple sentence level) were appropriate for the children in the present study.

The Childhood Autism Rating Scale (CARS) (Schopler, Reichler, & Renner, 1988) is a 15-item behavioral rating scale used to identify autism. Items are rated on a scale of 1 (normal) to 4 (severely abnormal). Research staff used the CARS to screen for the presence of autistic symptoms in the TD and DD groups.

For participants recruited by mail, all of our assessment instruments were parent questionnaires. They included:

Background Information Questionnaire: This instrument asked for demographic information about current age of child, race, ethnicity, parent education level, family composition, birth history, age of first parent concerns, symptoms causing first concerns, age(s) of any diagnoses, source(s) of any diagnoses, developmental regression, age of regression when applicable, known genetic or medical conditions, and family history of DD/ASD.

The *Ages and Stages Questionnaire* (ASQ) (Bricker & Squires, 1995) is a standardized screener for developmental concerns in children from 4 to 60 months of age. Nineteen different questionnaires cover designated age intervals. Each questionnaire covers fine motor, gross motor, problem solving, communication, and personal social areas of development. Each subsection is scored on a pass/fail basis.

The *Social Communication Questionnaire* (SCQ) (Rutter, Bailey & Lord, 2003) is a brief parent questionnaire available in two forms—Lifetime and Current. We used the Current form, which gives equivalent results to the Lifetime form for preschool children. The published cut-off score for Autistic Disorder on the SCQ is 15; however, Corsello, Cook, and Leventhal (2003) found that a cut-off score of 11 markedly increased the sensitivity of the SCQ for children under the age of 5 years (the age range of the majority of our subjects), while only slightly decreasing the specificity. In addition, we included children with a broader range of ASD in the current study, not only those with Autistic Disorder. For both reasons, we chose the lower SCQ cut-off.

Ascertainment

Participants from the collaborating study. The parents of 47 children in the collaborating study completed FYI-R questionnaires. We used data from the assessment instruments described above to ascertain children eligible for the current study. For children with ASD, additional inclusion criteria were a clinical diagnosis and meeting autism spectrum criteria on the ADOS and/or the ADI-R. Children in the DD group had to have a clinical diagnosis of a developmental disability and exhibit an overall cognitive delay of at least 2 standard deviations (*SD*) on the total MSEL or Leiter-R Brief IQ, or a delay of at least 1.5 *SD* in 2 or more developmental domains, including motor, nonverbal cognitive skills, expressive language, receptive language, and adaptive behavior. Exclusion criteria for both DD and ASD children included profound delay, severe cerebral palsy, vision or hearing impairments that were not corrected to within normal limits, or genetic disorders associated with ASD (e.g., fragile X, Rett syndrome). The diagnosis of ASD was ruled out for children in the DD group via careful observation over several hours of assessment, and results of the ADOS and CARS. Children in the TD group had to score within the normal range on developmental, cognitive, and adaptive measures. The possibility of an ASD was ruled out via observation and completion of the CARS. Of 47 potential participants, 43 children were eligible for inclusion, whereas 4 were excluded due to failure to meet the above criteria.

Participants recruited by mail. The parents of 61 children were recruited by mail. For these families, we required the children to be 5 years or younger, the appropriate age range for the ASQ. Inclusion criteria for the ASD participants included a reported clinical diagnosis of an ASD, and a score greater than 11 on the SCQ. We required that children with DD have a reported clinical diagnosis of DD, and fail at least one section of the ASQ other than, or in addition to, the fine or gross motor section. Children with DD also needed to have scores on the SCQ less than 11. Children were excluded from the study based on reported genetic diagnoses associated with ASD (e.g., fragile X, Rett syndrome). We included children in the TD group if their parents reported no history of developmental concerns and the child passed each section of the ASQ. Of 61 children recruited by mail, 11 failed to meet inclusion criteria, leaving 50 children from this subsample in the study.

Results

FYI-R Questionnaire and Scoring

Responses on the FYI-R were scored with a procedure derived from a normative sample of over 1300 participants (Reznick et al., in press). FYI questions are grouped into eight constructs. Each question can be designated as suggesting risk if a parent selects a

response that was highly unlikely in the normative sample. Risk scores are summed within each of the eight constructs. Because constructs differ on the number of relevant questions, each construct is converted to a common scale. Specifically, within each construct, a child is awarded a risk score between 0 and 50 on the basis of a quasi-logarithmic scale that assigns more risk points to children whose parents report more atypical observations. The distributions across questions vary somewhat, but in general, a risk-point total near the 50th percentile of the normative sample receives a risk score of 10, a risk-point total near the 75th percentile receives a risk score of 13, a risk-point total near the 90th percentile receives a risk score of 20, a risk-point total near the 95th percentile receives a risk score of 30, a risk-point total near the 98th percentile receives a risk score of 40, and a risk-point total above the 99th percentile receives a risk score of 50. The risk scores assigned to other risk-point totals are interpolated within this scale. The overall FYI risk score for a child is calculated by adding the risk scores across constructs and then dividing the total by 8 (the number of constructs). For example, if a child receives 20 risk points on the Imitation construct, 30 points on Reactivity, 30 points on Repetitive Behavior, and 0 points on the other five constructs, his/her overall FYI-R risk score would be 10 (i.e., 80 divided by 8).

Preliminary Analyses

We first examined several extraneous factors that could have influenced the risk scores. Children ranged in age between 14 and 75 months, with a median age of 41 months. There were no significant age differences among the three groups. Age was not correlated with the risk score for the combined sample or for any of the three groups. Finally, we divided the participants into an older and younger group on the basis of a median split and found no difference in risk score for these two groups.

We next compared risk scores between children with ASD who were recruited via our collaborating study and those with ASD whose parents participated by mail. The concern prompting this comparison was that we had more diagnostic information for the children in the collaborating study than for the latter group; however, an ANOVA revealed no significant difference in risk scores between these two subgroups.

Given the sex-linked prevalence of developmental disorders, it is not surprising that we recruited more males than females, particularly in the disability groups (see Table 1). An ANOVA across all three groups, however, revealed no main effect of sex or sex by group interaction on the risk score, but within groups, boys with autism had higher risk scores than girls with autism, F (1, 36) = 6.16, $p < .05$.

Analysis of Group Differences on Total Risk Score

Figure 1 shows the distribution of risk scores for the three groups. Inspection of Fig. 1 reveals that most of the participants in the ASD group were reported to have shown many atypical behaviors at 12 months. Even with an extremely strict risk criterion (99th percentile) that would identify only 1% of children in our normative sample, over half of the children with ASD would have been classified as at risk at

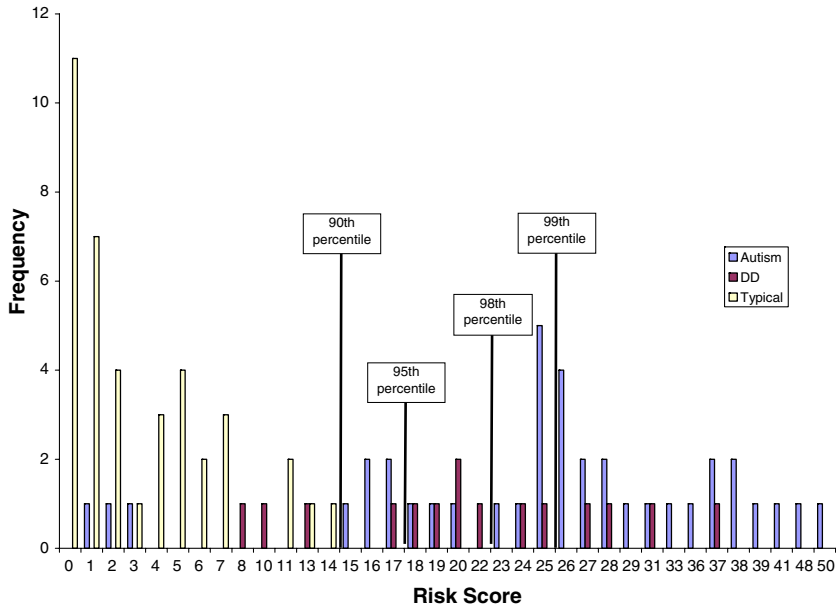

Fig. 1 Distribution of autism risk scores by group

12 months. Dropping down to the 90th percentile, 92% of the children with ASD would have been deemed at risk, but there is a precipitous drop in identification efficacy thereafter. Indeed, the three children with ASD who have the lowest FYI scores are dispersed among the scores for typical children.

The range of retrospective risk scores for children diagnosed with DD is notably lower than the range for children with ASD diagnoses. A few children with DD have risk scores above 30 and a few have scores in the typical range, but most risk scores for these children indicate considerable atypical behavior at 12 months. Finally, the children in the TD group have risk scores that reflect virtually no atypical behaviors.

We calculated an ANOVA comparing total risk scores for the three groups. To reduce heterogeneity of variance, we removed the three outlying scores in the ASD group. Each of these three children with ASD (two girls and one boy) scored well within the range of children in the TD group on the FYI-R, placing them a full standard deviation below their nearest neighbor with ASD; thus, we deemed it inappropriate to include their scores in the ASD group mean. Given the various onset patterns for autism symptoms, it is not surprising to find children with ASD whose parents report typical behavior at 12 months.

As apparent in Fig. 1, the effect of group is quite strong, F (2, 87) = 127.96, $p < .01$. The omega squared associated with this ANOVA is .74, a large effect size (Kotrlik & Williams, 2003). Post hoc analysis using the Tukey's Studentized Range (HSD) showed significant differences between the ASD group and both the DD and TD groups at an alpha level of .05. In addition, the DD group had a significantly higher mean risk score than the TD group. The means for the ASD, DD, and TD groups were 28.04 (SD = 8.8, range 14.9–49.5), 21.21 (SD = 7.76, range 8.4–36.6), and 3.28 (SD = 3.85, range 0–14.4), respectively.

Item Analysis

In Table 2 we present results from the ANOVAs for the individual items associated with each of the eight constructs on the FYI-R. Results of the post-hoc analyses using HSD, and the between group comparisons are presented in the three right-hand columns in Table 3. These analyses revealed a main effect for group for each of the individual items in the four constructs in the Social-Communication domain. As indicated in Table 2, individual items related to Social Orienting and Receptive Communication and to Social Affective Engagement were particularly useful for differentiating between the ASD and DD groups. On the other hand, items related to Imitation and to Expressive Communication generally differentiated children with ASD or DD from children in the TD group, but children in the ASD and DD groups performed similarly on most items on those two constructs.

Among the four constructs in the Sensory-Regulatory domain, ANOVAs revealed a main effect of group for 16 of 24 items. The scores on 14 of these items differentiated the ASD group from the TD group, and scores on 5 items differentiated the ASD group from the DD group.

Analysis of Constructs

The eight constructs were analyzed individually using ANOVA procedures. As shown in Table 3, each ANOVA resulted in a significant F statistic ($p < .01$) with the exception of the analysis of Regulatory Patterns. Post-hoc comparisons of the groups revealed higher mean scores for the ASD group than the DD and TD groups on the constructs of Social Orienting and Receptive Communication, Social Affective Engagement, and Reactivity. The DD and TD groups did not differ significantly from each other on these three constructs. The ASD and DD groups had higher mean scores than the TD group on Imitation, Expressive Communication, Sensory Processing, and Repetitive Play and Behavior; however, on these four constructs, the ASD and DD groups did not differ significantly from each other.

Analyses of Domains

ANOVAs demonstrated significant differences among the three groups in Social-Communication domain, F (2, 87) = 173.5, $p < .0001$ and Sensory-Regulatory domain, F (2, 87) = 22.3, $p < .0001$. An HSD post hoc analysis revealed that the ASD group mean (37.20, SD = 11.6, range 14.8–50) for the Social-Communication domain was significantly different from both the DD (28.0, SD = 7.8, range 14.5–42.8) and TD (2.2, SD = 3.7, range 0–13.8) groups. In addition, the DD and TD groups were significantly different from each other. For the Sensory-Regulatory domain, the post hoc analysis revealed a significant difference between the means for the ASD group (18.8, SD = 11.8, range 2.3–49.0) and the TD group (4.3, SD = 5.9, range 0–26.5), but the ASD group did not differ significantly from the DD group (14.4, SD = 11.5, range 2.3–44.3).

Table 2 Item-by-Item results

Item	Sig	Aut v Typ	Aut v DD	DD v Typ
Social orienting & receptive communication				
1. Did your baby turn or look at you when you called baby's name?	<.0001	*	*	
5. Did your baby seem to have trouble hearing?	<.0001	*	*	
10. When you pointed to something interesting, did your baby turn to look at it?	<.0001	*	*	*
12. Did your baby look at people when they began talking, even when they were not talking directly to your baby?	<.0001	*	*	
14. Did your baby look up from playing with a favorite toy if you showed him or her a different toy?	<.0001	*	*	
31. Did your baby seem interested in other babies his or her age?	<.0001	*	*	*
35. When you said ''Where's (a familiar person or object)'' without pointing or showing, would your baby look at the person or object named?	<.0001	*	*	*
50. What did you typically have to do to get your baby to look up from playing with a favorite toy?	<.0001	*		*
52. What did you typically have to do to get your baby to turn towards you?	<.0001	*	*	*
Social affective engagement				
4. During familiar games like ''I'm gonna get you,'' did your baby get excited because he or she knew what would happen next?	<.0006	*		
7. In new or strange situations, did your baby look at your face for comfort?	<.0001	*		
16. Was it easy to understand your baby's facial expressions?	<.0001	*	*	
18. Did your baby smile while looking at you?	<.0001	*	*	
19. Did your baby try to get your attention to show you something interesting?	<.0001	*	*	
20. Did your baby try to get your attention to play games like peek-a-boo?	<.0001	*	*	
21. Did your baby try to get your attention to obtain a favorite toy or food?	<.0001	*	*	
22. Did your baby try to get your attention to play physical games, like swinging, tickling, or being tossed in the air?	<.0001	*	*	
Imitation				
24. Did your baby copy or imitate you when you made sounds or noises with your mouth?	<.0001	*		*
25. Did your baby copy or imitate your actions, like sticking out your tongue, clapping your hands, or shaking your head?	<.0001	*		*
26. Did your baby copy or imitate you when you did something with a toy or object, like shaking a rattle or banging a spoon on the table?	<.0001	*	*	*
49. When you introduced your baby to a new game (peek-a-boo, so-big, patty-cake, etc) how did your baby respond?	<.0001	*		*
53. What did you typically have to do to get your baby to smile or laugh at you?	<.0001	*	*	
58. If you started a game by copying or imitating a sound your baby made, what did your baby typically do?	<.0001	*		*
Expressive communication				
29. Did your baby try to get your attention by making sounds and looking at you at the same time?	<.0001	*	*	*
32. Did your baby babble by putting sounds together, such as 'ba-ba', ga-ga', or 'ba-dee'?	<.0001	*		*
34. Did your baby use gestures such as raising arms to be picked up, shaking head, or waving bye-bye?	<.0001	*		*
38. Did your baby communicate with you by using his or her finger to point at objects or pictures?	<.0001	*		*
Sensory processing				
3. Did your baby seem overly sensitive to your touch (for example, fuss or pull away when you touch him or her)?	.005	*		
6. When you and your baby were facing each other, did your baby turn his or her eyes to avoid looking at you?	<.0001	*	*	
9. Did your baby spit out certain textures of foods, such as lumpy or chunky pieces?	n.s.			
17. Did your baby forcefully press his or her face, head, or body against people or furniture?	.004	*		
23. When your baby was awake and you picked him or her up, did your baby's body feel loose or floppy?	<.0001		*	*
59. When your baby was awake and not eating, did your baby keep a toy or object in his mouth?	n.s.			

Table 2 continued

Item	Sig	Aut v Typ	Aut v DD	DD v Typ
Regulatory patterns				
28. Were your baby's sleeping and waking patterns regular from day to day?	n.s.			
41. Were your baby's feeding patterns regular from day to day?	n.s.			
54. On a typical night, how many hours did your baby sleep?	.009	*		
55. On a typical night, how many times did your baby wake up?	.013	*		
Reactivity				
15. Did your baby get upset when you needed to switch your baby from one activity to another one?	n.s.			
27. Was it difficult to calm your baby once he or she became upset?	.001	*	*	
57. Which of the following best described your baby's typical day? (# of times the baby got upset during the day)	n.s.			
Repetitive behavior				
11. Was your baby content to play alone for an hour or more at a time?	<.0001	*		
13. Did your baby rock his or her body back and forth over and over?	.01			*
30. Did your baby get stuck doing a simple activity over and over?	<.0001	*	*	
33. Did your baby enjoy staring at a bright light for long periods of time?	.004	*		
37. Did your baby seem to get stuck on playing with a part of a toy (such as an eyeball, label, wheel or tag), instead of the whole toy?	<.0001	*	*	
42. Did your baby enjoy rubbing or scratching toys or objects for long periods of time?	n.s.			
43. Did your baby seem to get his or her body stuck in a position or posture that was hard to move out of?	.003	*		
44. Did your baby enjoy making objects spin over and over in the same way?	.025	*		
45. While lying down, did your baby enjoy kicking his or her feet over and over for long periods of time?	<.0001	*		*
46. Did your baby stare at his or her fingers while wiggling them in front of his or her eyes?	n.s.			
48. Was your child interested in a variety of toys throughout a typical day?	<.0001	*		*

Identification of Children Who Are at Risk for ASD

Figure 1 illustrates the implications of using different cut-off points for total risk scores in identifying children at risk for ASD. Using a minimum score of 15 as the cut-off for ASD risk (i.e., scores at or above the 90th percentile in our normative sample), 35 of 38 children with ASD would be identified, as would 12 of 15 children with DD, and 0 of 40 children with TD. Applying this cut-off to our current sample, the FYI-R sensitivity is .92, specificity is .78, positive predictive value is .74 and negative predictive value is .93. If we increase the cut-off to require total risk scores of greater than 22 (i.e., scores ≥ the 98th percentile for our normative sample), we would identify 27 of 38 children with ASD, 6 of 15 children with DD and again 0 of 40 children with TD. Using this cut-off score, the FYI-R sensitivity is .71, specificity is .89, positive predictive value is .82, and negative predictive value is .82.

Table 3 Analyses of variance for FYI-R construct scores within two domains

Domain/Constructs	df	*F*	MSE	*P*
Social-communication domain				
Social Orienting and Receptive Communication [a, b]	2	130.83	(113.4)	<.0001
Social Affective Engagement [a, b]	2	49.8	(158.5)	<.0001
Imitation [b, c]	2	92.1	(153.6)	<.0001
Expressive Communication [b, c]	2	116.1	(117.4)	<.0001
Sensory-regulatory domain				
Sensory Processing [b, c]	2	12.0	(191.1)	<.0001
Regulatory Patterns	2	3.8	(286.4)	0.027
Reactivity [a, b]	2	7.8	(301.9)	0.0008
Repetitive Play & Behavior [b, c]	2	25.0	(146.7)	<.0001
S within- group error	87			

[a] ASD > DD [b] ASD > TD [c] DD > TD ($p < .05$; post hoc [HSD] comparisons of group means)

Within-Group Analysis for the ASD Group

We also examined the possibility of significant relationships between the autism risk score and several variables specific to the ASD group. The variables of interest included age of initial diagnosis of ASD, and age of first parental concerns, both assessed by parent report, and severity of autism symptoms, assessed by the ADI-R or the SCQ. The three children with ASD who had outlying scores in the low end of the TD range were excluded from these analyses, and we had missing data from some other participants as well (varying from analysis to analysis, as indicated by the degrees of freedom reported below). A Pearson product moment correlation indicated moderately low, nonsignificant relations between the autism risk score and age of initial ASD diagnosis, r (29) = –0.30, p = .113, as well as between autism risk scores and the child's age when parents were first concerned, r (32) = –0.32, p = .071.

To examine the relation between FYI-R risk score and autism symptom severity at preschool age, we divided our participants into four subgroups: verbal and nonverbal children with ADI-R scores, and verbal and nonverbal children with SCQ scores. We subgrouped children in this way due our use of different instruments to measure severity of symptoms (ADI-R vs. SCQ), and because both of these instruments omit some items for nonverbal children (i.e., children with expressive language no higher than the phrase level). For each subgroup, we calculated a mean and SD for the ADI-R or SCQ total score, and computed a z-score for each child in the subgroup. The z-scores were then combined across groups to yield one commensurate variable reflecting symptom severity among all children in the ASD group. There was a significant relationship between the autism risk score and autism symptom severity r (35) = .54, p = .0008.

Previous research suggests that a significant proportion of children with ASD experience developmental regression after 12 months of age. In the current study, 14 caregivers reported regression in their children with ASD, with 12 of 14 reporting regression after 12 months of age. For our purposes, the regression noted by caregivers could be in any developmental area, based on information provided in the background information questionnaire or during the ADI-R interview. Because regression is overwhelmingly reported after 12 months, we would expect these children to be perceived as being at less risk at 12 months. An analysis of group means for autism risk score confirmed this prediction. Participants in the ASD group with reported regression had a mean risk score of 21.6 (SD = 10.6, range 1.0–37.9), whereas those without regression had a mean risk score of 29.6 (SD = 9.8, range 14.9–34.6), [t *(31) = 2.07,* $p < .05$]. This difference is strongly affected by the three children with the extremely low risk scores, all of whom were reported to have regressed. If these three participants are removed from the analysis, the mean autism risk scores of children with and without reported regression are not significantly different. This finding suggests that most children with reported regression nevertheless already show considerable risk for ASD at 12 months.

Discussion

The current study contributes to efforts to develop a screening instrument with utility for identifying 12-month-olds in the general population at risk for an eventual diagnosis of ASD. The present data confirm that most retrospectively reported items on the FYI are sensitive to clinically relevant group differences among preschool children. Thus, we have support for assuming that the FYI can measure distinctive characteristics of 12-month-olds who will later be diagnosed with ASD.

Second, all four Social-Communicative constructs and three of four Sensory-Regulatory constructs reflected group differences among ASD, DD, and TD groups. Combined with the information regarding the performance of the children on individual items, these findings contribute to our understanding of the phenotypic features of infants later diagnosed with ASD and other DD. For example, the present data suggest that high risk scores on Social Orienting and Receptive Communication and on Social Affective Engagement are rare among children with DD but common among children with ASD. On the other hand, many children with DD as well as those with ASD had relatively high risk scores on Imitation and on Expressive Communication. Analyses of the discriminative value of individual items and the profile of risk scores across constructs can lead to improvements in early assessment and intervention.

We found a large effect size for group differences in mean total risk scores. Applied to the task of classifying children as at risk or not at risk for ASD, the FYI-R performed relatively well. Various cut-off scores (relative to the normative sample) can be used to optimize the FYI's performance. For example, setting the threshold at the 90th percentile provides very high sensitivity and negative predictive value, but somewhat lower specificity and positive predictive value. Increasing the threshold to the 98th percentile raises the specificity and positive predictive value of

the instrument, but lowers the sensitivity and negative predictive value. Thus, the current study provides a basis for comparing alternate thresholds for FYI scores and emphasizes the fact that the optimal threshold will vary relative to one's purpose. For example, setting a high threshold for autism risk could identify children quite likely to be diagnosed with ASD, and thus would be useful for research that depends on accurate identification of participants with ASD. The positive correlation between FYI-R risk score and severity of autism symptoms at preschool age suggests that the FYI may have particular utility for identifying infants at risk for more severe manifestations of ASD. However, the present data demonstrate that the cost of achieving a high level of specificity with the FYI would be to miss a substantial proportion of children with ASD (30% in our sample at the 98 percentile). This cost may be acceptable in some research contexts, but would be unacceptable in epidemiological research or large-scale public health screenings aimed at identifying infants at risk for ASD. In the latter situations, a lower threshold may have more utility.

The current data provide valuable information on infant characteristics of children later diagnosed with ASD, and will guide our efforts as we refine the FYI. We also acknowledge the challenges and limitations of this research. The study included a relatively small sample, and it was especially limited in participants with DD. Our data, as well as data published by others, suggest that the challenge of distinguishing between ASD and other DD among young children is much greater than the challenge of distinguishing between ASD and TD (e.g., Dietz et al, 2006; Robins et al., 2001). We join our colleagues (e.g., Dosreis, Weiner, Johnson, & Newschaffer, 2006; Robins & Dumont-Mathieu, 2006) in asserting that an instrument that over-identifies risk for ASD among children with other DD is more beneficial clinically than an instrument that misses large numbers of children truly at risk for ASD.

In the current study, the Sensory-Regulatory domain risk scores contributed to the identification of children with ASD, but were lower overall than expected. We cannot determine whether many infants at risk for ASD fail to show extreme Sensory-Regulatory symptoms, or whether these symptoms are less salient to parents and/or harder to recall in retrospect than Social-Communication domain symptoms. This will be a subject for future research.

Another issue related to screening for ASD risk at 12 months is that children who show a pattern of typical development followed by regression after 12 months will likely be missed in the screening. This was the case among our sample for three children with ASD. We accept the possibility that some infants eventually diagnosed with ASD will look typical to both parents and professionals at 12 months. This would set a natural limit on the FYI's sensitivity. However, the current study and previous research (Baranek, 1999) indicate that many children with ASD and reported regression are exhibiting symptoms in infancy that distinguish them from their peers with TD or DD, even though parents may not be conscious of these features. Thus, a successful prospective screening requires an efficient and effective way of identifying subtle characteristics that may be hard for a parent to evaluate relative to normal individual differences in children.

A specific limitation of our study was that more than half of our sample participated entirely through responses to written questionnaires. This prevented us from directly characterizing a large proportion of the children and investigating questions that depend on direct observation. For example, we would be interested in knowing whether children who receive diagnoses of Autistic Disorder have higher FYI-R scores than children who fall in the broader autism spectrum but do not meet strict criteria for Autistic Disorder. Due to the acknowledged lack of agreement among clinicians in differentiating disorders within the autism spectrum, the large number of clinicians across varied professions providing diagnoses for our participants, and the lack of in-depth diagnostic information (ADI-R; ADOS) on a portion of the sample, we were not able to examine the relation between FYI-R scores and subtypes of ASD in the current investigation.

There are risks inherent in trying to apply retrospective information from parents of preschoolers to the task of crafting a tool to prospectively screen 12-month-olds for risk of ASD. The retrospective data undoubtedly vary from data we would get from reports actually completed at 12 months in ways we cannot fully predict; thus, we will be cautious in using the FYI-R data to refine the FYI, and we do not assume that the FYI will demonstrate similar test characteristics in a prospective study as the FYI-R did in this retrospective study (i.e., sensitivity, specificity, positive and negative predictive values).

Finally, the identification of risk for ASD in 12-month-olds raises some difficult issues regarding public health policies and family-centered services. Although further clinical assessment for any child failing a valid developmental screening is warranted, there is no "gold standard" that allows for immediate confirmation that a 12-month-old with a FYI score in the risk range indeed is on the trajectory for developing ASD.

Conducting screenings for ASD at such a young age entails grave responsibilities for researchers and practitioners regarding the information they present to families, and the care they take in doing so. These responsibilities are heightened by the fact that broad-based early screening for a specific developmental disorder such as ASD inevitably will lead to some false positives among children who are completely healthy and typically developing. Clark and Harrington (1999) have discussed this and other clinical issues related to screening for disorders with low base rates in the general population. In our view, researchers and clinicians must consider the ethical ramifications of early screening for ASD and similar disorders. Appropriate accountability includes demonstrating the reliability and validity of screening instruments, and studying the relative costs versus benefits of earlier versus later identification of children at risk for ASD. Researchers and clinicians alike have a responsibility for understanding the limits of available screening instruments, and communicating with caregivers honestly and respectfully about the limits of early screening as well as the present state of our knowledge regarding early intervention.

Acknowledgments We are grateful to the families that participated in this study. We also thank the staff of the Sensory Experiences Project for their assistance with data collection. The authors acknowledge partial funding support from the National Institute for Child Health and Human Development (NICHD R01-HD42168), the National Alliance for Autism Research (#681), and a Smith Graduate Research Grant from the University of North Carolina at Chapel Hill. In addition, we deeply appreciate the contributions of the parents who participated in the study.

References

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders*, (4th ed.) Washington, DC: American Psychiatric Association.

Baird, G., Charman, T., Baron-Cohen, S., Cox, A., Swettenham, J., Wheelwright, S., Drew, A (2000). A screening instrument for autism at 18 months of age: A 6-year follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*(6), 694–702.

Baranek, G. T. (1999). Autism during infancy: A retrospective video analysis of sensory-motor and social behaviors at 9–12 months of age. *Journal of Autism and Developmental Disorders, 29*, 213–224.

Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *British Journal of Psychiatry, 161*, 839–843.

Baron-Cohen, S., Cox, A., Baird, G., Swettenham, J., Nightingale, N., Morgan, K., Drew, A., Charman, T. (1996). Psychological markers in the detection of autism in infancy in a large population. *British Journal of Psychiatry, 168*, 158–163.

Bricker, D., Squires, J. (1995). *Ages and Stages Questionnaire: A parent-completed child monitoring system* (2nd ed.) Baltimore, MD: Brookes.

Centers for Disease Control and Prevention. (2006). Mental health in the United States: parental report of diagnosed autism in children aged 4–17 years-United States, 2003–2004. *Morbidity and Mortality Weekly Report,* 55(17), 481–486.

Clark, A., & Harrington, R. (1999). On diagnosing rare disorders rarely: Appropriate use of screening instruments. *Journal of Child Psychology and Psychiatry, 40*, 287–290.

Corsello, C., Cook, E., & Leventhal, B. (2003). *A screening instrument for autistic spectrum disorders*. Presented at the Biennial Meeting of the Society for Research in Child Development, Miami, FL.

Dietz, C., Swinkels, S., van Daalen, E., van Engeland, H. & Buitelaar, J. K. (2006). Screening for autistic spectrum disorder in children aged 14–15 months. II: Population screening with the Early Screening of Autistic Traits Questionnaire (ESAT). Design and general findings. *Journal of Autism and Developmental Disorders, 36*, 713–722.

Dosreis, S., Weiner, C. L., Johnson, L., & Newschaffer, C. J. (2006). Autism spectrum disorder screening and management practices among general pediatric providers. *Journal of Behavioral and Developmental Pediatrics, 27*(Supplement 2), S88–S94.

Kotrlik, J. W., & Williams, H. A. (2003). The incorporation of effect size in information technology, learning, and performance research. *Information Technology, Learning, and Performance Journal, 21*, 1–7.

Landa, R., & Garrett-Mayer, E. (2006). Development in infants with autism spectrum disorders: a prospective study. *Journal of Child Psychology and Psychiatry, 47*, 629–638.

Lord, C., Rutter, M., Dilavore, P., & Risi, S. (1999). *The Autism Diagnostic Observation Schedule (ADOS)*. Los Angeles: Western Psychological Corporation.

Maestro, S., Casella, C., Milone, A., Muratori, F., & Palacio-Espasa, F. (1999). Study of the onset of autism through home movies. *Psychopathology, 32*, 292–300.

Mullen, E. M. (1995). *Mullen scales of early learning* (AGS ed.). Los Angeles: Western Psychological.

Osterling, J. G., Dawson, G., & Munson, J. A. (2002). Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Development & Psychopathology, 14*, 239–251.

Reznick, J. S., Baranek, G. T., Reavis, S., Watson, L. R., & Crais, E. R. (in press). A parent-report instrument for identifying one-year-olds at risk for an eventual diagnosis of autism: The First Year Inventory. *Journal of Autism and Developmental Disorders*.

Robins, D. L., & Dumont-Mathieu, T. M. (2006). Early screening for autism spectrum disorders: Update on the Modified Checklist for Autism in Toddlers and other measures. *Journal of Developmental and Behavioral Pediatrics, 27*(Supplement 2), S111–S119.

Robins, D. L., Fein, D., Barton, M. L., Greene, J. A. (2001). The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 31*, 131–144.

Roid, G. H., & Miller, L. J. (1997). *Leiter international performance scale revised*. Wood Dale, IL: Stoelting Co.

Rutter, M., Bailey, A., & Lord, C. (2003). *Social Communication Questionnaire*. Los Angeles, CA: Western Psychological Services.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism Diagnostic Interview-Revised (ADI-R)*. Los Angeles, CA: Western Psychological Services.

Schopler, E., Reichler, R. J., & Renner, B. R. (1988). *The Childhood Autism Rating Scale*. Los Angeles, CA: Western Psychological Services.

Siegel, B. (2004). *Pervasive Developmental Disorders Screening Test-II (PDDST-II): Early childhood screener for autistic spectrum disorders*. San Antonio, TX: Harcourt Assessment.

Swinkels, S. H. N., Dietz, C., van Daalen, E., Kerkhof, I. H. G. M., van Engeland, H., & Buitelaar, J. K. (2006). Screening for autistic spectrum in children aged 14 to 15 months. I: The development of the Early Screening of Autistic Traits Questionnaire (ESAT). *Journal of Autism and Developmental Disorders, 36*, 723–732.

Sparrow, S., Balla, D., & Cicchetti, D. (1984). *Vineland adaptive behavior scales*. Circle Pines, MN: American Guidance Service.

Yirmiya, N., Gamliel, I., Pilowsky, T., Feldman, R., Baron-Cohen, S., & Sigman, M. (2006). The development of siblings of children with autism at 4 and 14 months: social engagement, communication, and cognition. *Journal of Child Psychology and Psychiatry, 47*, 511–523.

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience, 23*(2–3), 143–152.