*Original Article*

autism

Autism
1–10

DOI: 10.1177/1362361315593941
aut.sagepub.com
SAGE

# Factor structure and psychometric properties of the revised Home Situations Questionnaire for autism spectrum disorder: The Home Situations Questionnaire-Autism Spectrum Disorder

**Monali Chowdhury[1], Michael G Aman[2], Luc Lecavalier[2], Tristram Smith[3], Cynthia Johnson[4], Naomi Swiezy[5], James T McCracken[6], Bryan King[7], Christopher J McDougle[8], Karen Bearss[9], Yanhong Deng[10] and Lawrence Scahill[9]**

**Abstract**

Previously, we adapted the Home Situations Questionnaire to measure behavioral non-compliance in everyday settings in children with pervasive developmental disorders. In this study, we further revised this instrument for use in autism spectrum disorder and examined its psychometric properties (referred to as the Home Situations Questionnaire-Autism Spectrum Disorder). To cover a broader range of situations and improve reliability, we prepared seven new items describing situations in which children with autism spectrum disorder might display non-compliance. Parents completed ratings of 242 children with autism spectrum disorder with accompanying disruptive behaviors (ages 4–14 years) participating in one of two randomized clinical trials. Results from an exploratory factor analysis indicated that the Home Situations Questionnaire-Autism Spectrum Disorder consists of two 12-item factors: Socially Inflexible ($\alpha = 0.84$) and Demand Specific ($\alpha = 0.89$). One-to-two-week test–retest reliability was statistically significant for all scored items and also for subscale totals. The pattern of correspondence between the Home Situations Questionnaire-Autism Spectrum Disorder and parent-rated problem behavior, clinician-rated repetitive behavior, adaptive behavior, and IQ provided evidence for concurrent and divergent validity of the Home Situations Questionnaire-Autism Spectrum Disorder. Overall, the results suggest that the Home Situations Questionnaire-Autism Spectrum Disorder is an adequate measure for assessing non-compliance in a variety of situations in this population, and use of its two subscales will likely provide a more refined interpretation of ratings.

**Keywords**

non-compliance, psychometric properties

## Introduction

The original Home Situations Questionnaire (HSQ; Barkley and Edelbrock, 1987) is a 16-item scale developed to examine the severity of non-compliant behavior in children with disruptive behavior disorders. The items describe different situations in which children commonly display non-compliant behavior. Parents are asked to indicate whether the child has a problem with compliance in each situation and, if so, to rate the severity of non-compliance

[1]Allegheny College, USA
[2]The Ohio State University, USA
[3]University of Rochester Medical Center, USA
[4]University of Florida, USA
[5]Indiana University, USA
[6]University of California, Los Angeles, USA
[7]University of Washington, USA
[8]MassGeneral Hospital for Children, USA
[9]Emory University, USA
[10]Yale Center for Analytical Science, USA

**Corresponding author:**
Michael Aman, The Ohio State University, McCampbell Hall, 1581 Dodd Drive, Columbus, OH 43210, USA.
Email: aman.1@osu.edu; michael.aman@osumc.edu

on a 1–9 Likert scale. Higher scores indicate greater non-compliance. This study explored the psychometric properties of an expanded version of the HSQ modified for children with autism spectrum disorders (HSQ-ASD).

Three groups of investigators have conducted factor analyses of previous versions of the HSQ. Breen and Altepeter (1991) reported a four-factor structure (non-family transactions, custodial transactions, task performance transactions, and isolate play) in a sample of 995 non-referred children, ages 4–11 years. In contrast, DuPaul and Barkley (1992) studied the structure of a revised 14-item version of the HSQ in 625 non-referred children (ages 6–12 years) and found only one factor. Adams et al. (1995) modified the HSQ for use with adolescents and developed the adolescent HSQ-parent report form (AHSQ-pr). They found a four-factor solution with this modified instrument in 943 non-referred adolescents (ages 11–17 years). The first factor comprised situations that involved non-family members; however, the remaining three factors did not appear to assess clinically meaningful constructs. Overall, results from previous factor analytic studies of the HSQ do not provide a consistent picture of the underlying components of situational non-compliance. However, these studies were conducted in non-clinical samples using different versions of the instrument. Thus, it is not clear that these findings are relevant to children with whom we work, namely those with developmental disabilities such as autism spectrum disorder (ASD).

In preparation for a randomized clinical trial (RCT) comparing risperidone alone to risperidone plus parent training (PT), the Research Unit on Pediatric Psychopharmacology (RUPP) Autism Network investigators modified the HSQ to measure outcome for children with pervasive developmental disorders (PDDs) (Aman et al., 2009; Scahill et al., 2009). The RUPP investigators added five situations to the HSQ in order to make the instrument more relevant to children with ASD (Chowdhury et al., 2010). Using then-current terminology from *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; DSM-IV-TR), we called this instrument the Home Situations Questionnaire-Pervasive Developmental Disorders (HSQ-PDD) version. The psychometric properties of the HSQ-PDD were evaluated in the 124 children who participated in the 24-week clinical trial. To our knowledge, this was the first psychometric study of the HSQ in participants with ASD. The total score on the HSQ-PDD (expressed as per-item mean) was sensitive to change and showed that combined treatment was superior to medication alone (Aman et al., 2009). Exploratory factor analyses (EFAs) revealed a two-factor structure solution that seemed clinically and empirically useful for characterizing non-compliance in ASD. Based on the content of the items, these factors were labeled as Socially Inflexible (14 items) and Demand Specific (6 items). Four items were discarded because they did not load on either factor. Subsequent analyses of associations among the Socially Inflexible and Demand-Specific subscales and a host of criterion measures provided evidence of adequate convergent and divergent validity (see Chowdhury et al., 2010). These analyses suggested that the modified HSQ for children with PDDs is an appropriate measure to assess behavioral non-compliance in children with PDDs.

The HSQ-PDD has been used to complement other measures of outcome in ASD by the RUPP Autism Network. In one RCT (Aman et al., 2009), it was the primary outcome measure. RUPP investigators have used it in other RCTs involving children with ASD (e.g. a trial of guanfacine (https://clinicaltrials.gov/ct2/show/NCT01238575?term=autism+guanfacine&rank=1) and a follow-up investigation of children in the earlier PT RCT (Arnold et al., 2012)). At the time of this writing, our group has used the HSQ as one of the primary outcomes to assess the impact of PT in preschoolers with ASD and severe behavior problems (https://clinicaltrials.gov/ct2/show/NCT01233414?term=parent+training+autism&rank=5). The HSQ is becoming an outcome measure of increasing relevance in the ASD research landscape.

Further revisions of the HSQ-PDD are warranted for two reasons. First, items in the HSQ-PDD are unevenly distributed between the two HSQ subscales (14 for Socially Inflexible and only six for Demand Specific). Although the subscales seemed to capture related but somewhat different constructs, the total score on the HSQ-PDD depended heavily on non-compliance in *social* situations. In an effort to add to the Demand-Specific subscale (and consequently allow for a broader range of scores), seven new items thought to reflect Demand-Specific situations were added (these items are described in the Method and appear in Table 2). Second, in keeping with the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed., DSM-V; American Psychiatric Association (APA), 2013), we named the 27-item version as the HSQ-ASD.

The purpose of this investigation was to examine the psychometric properties of the revised and expanded 27-item version of the HSQ-ASD in a sample of children with ASDs with accompanying disruptive behaviors. Given the specific nature of our sample, results may *not* be generalizable to children without ASDs. For our statistical analysis, we predicted that the seven new items would factor selectively on the Demand-Specific subscale. We also predicted clear evidence of convergent and divergent validity with measures of problem behavior, adaptive behavior, and IQ, as well as high test–retest reliability.

## Method

### *Sample characteristics*

The participants were children, ages 3–14 years, with ASD and chosen for the presence of disruptive behaviors or attention deficit hyperactivity disorder (ADHD)

symptoms. All participants were drawn from two RCTs that were recently completed. The first was the Research Units on Behavioral Intervention (RUBI) study, which was a 24-week clinical trial of 180 preschool children who were randomly assigned to PT or parent education (PE) (Bearss et al., 2015). Entry criteria included (a) presence of ASD (autistic disorder, Asperger's disorder, or pervasive developmental disorder-not otherwise specified (PDD-NOS) using DSM-IV-TR), (b) ages 3–6 years and 11 months, (c) baseline score ⩾15 on the Irritability subscale of the Aberrant Behavior Checklist (ABC), (d) Clinical Global Impressions Severity (CGI-S) score of ⩾4 (indicating at least moderate severity of disruptive behavior), and (e) taking no psychotropic medication or stable medication for the last 6 weeks. We excluded children with Rett's Disorder or Childhood Disintegrative Disorder, a known serious medical condition that could interfere with child's ability to participate in the study, presence of a psychiatric disorder (psychotic disorder and major depression) that would require another treatment, current or past participation in a structured PT program, or receptive language <18 months of age.

The second RCT was a double-blind, parallel groups comparison of placebo and extended-release guanfacine in 62 children with ASD accompanied by hyperactivity, impulsiveness, and distractibility. Inclusion criteria were (a) presence of any ASD subtype (using DSM-IV-TR criteria), (b) ages 5 years through 14 years, inclusive, (c) baseline score ⩾25 on Hyperactivity subscale of the ABC, (d) CGI-S score ⩾4 for ADHD symptoms, and (e) free of psychotropic medications. We excluded children with IQ < 35, a significant medical condition, or a *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; DSM-IV) diagnosis of major depression, psychotic disorder, or bipolar disorder.

The pre-treatment assessment was virtually identical across these two trials. Merging of data from the two trials increased the sample size and expanded the age range and clinical characteristics of the sample. Instruments that were common to both samples included the HSQ, ABC, Children's Yale–Brown Obsessive-Compulsive Scale (CY-BOCS)-PDD, Vineland Adaptive Behavior Scales II, Child Symptom Inventory, and IQ, all of which are described in detail below. Both samples were assessed on the following additional instruments: (a) Autism Diagnostic Observational Schedule (Lord et al., 2000), a tool for confirming the presence of ASD; (b) the CGI-S subscale (Guy, 1976), a scale for gauging overall level of psychopathology; (c) Parent-Nominated Target Problems (Arnold et al., 2003), a method for tracking and assessing two of the most troublesome child problems, nominated by parents; (d) standard demographics forms; (e) medical and psychiatric history; and (f) concomitant treatment form, for tracking changes in medications and psychosocial/educational interventions. Thus, we knew a great deal about our participants based on this standardized battery of assessments.

The demographics of the RUBI and Guanfacine samples are described in Table 1. Mean age for the RUBI group was 4.7 and 8.4 years for the Guanfacine group. More than 85% of participants in both studies were male. Most had a diagnosis of Autistic Disorder, and most had IQs ⩾ 70. In total, 50% of RUBI participants and 39% of Guanfacine participants attended special education classes or schools. In terms of clinical severity, most children in both studies had Clinical Global Impression Severity scores of 5 (markedly ill) or 4 (moderately ill). The majority of both samples were white (87% for RUBI; 65% for Guanfacine), with African American participants comprising 8% of the RUBI sample and 17.7% of the Guanfacine group. Further details appear in Table 1; the right-most column contains demographic features for the combined sample.

## Study instruments

*HSQ.* As noted above, the 20-item HSQ-PDD included 14 Socially Inflexible subscale items and 6 Demand-Specific subscale items. Five of the current authors (M.C., M.G.A., L.L., K.B., and L.S.) added seven new items intended to capture Demand-Specific situations for preschool and school-age children. These items were short listed from a longer list of potential items through discussion and consensus among these authors based on their clinical experience with children in this population. The new items are italicized in Table 2. Examples include "When told to brush teeth" and "When asked to put clothes in proper places." Thus, the new HSQ (dubbed the HSQ-ASD) comprised 27 items. As in previous versions of the HSQ, parents were asked whether the given situation posed a problem for the child (yes/no) in the past 4 weeks. If the answer was yes, the parent then rated the severity on a scale ranging from 1 to 9. Thus, the possible range of scores for the HSQ was 0 (resulting from a "no" response) to 9. The total severity score is divided by the total number of items to derive a per-item mean.

*ABC.* The ABC is a reliable and valid standardized 58-item rating scale for assessing problem behaviors in individuals with developmental disabilities (Aman et al., 1985a, 1985b). It was empirically derived by principal components analysis (Aman et al., 1985b), and its original component structure has since been shown, by factor analysis, to be consistent in children with ASD (Kaat et al., 2014). The ABC subscales have been designated as follows: Irritability (15 items), Social Withdrawal (16 items), Stereotypic Behavior (7 items), Hyperactivity/Non-compliance (16 items), and Inappropriate Speech (4 items). Each item is rated on a 4-point Likert Scale from 0 (not a problem) through 3 (the problem is severe). Internal consistency,

**Table 1.** Demographic characteristics of RUBI and Guanfacine study samples.

| | RUBI | | Guanfacine | | RUBI + Guanfacine | |
|---|---|---|---|---|---|---|
| | *M*/*N* | SD/% | *M*/*N* | SD/% | *M*/*N* | SD/% |
| Age (years) | 4.72 | 1.06 | 8.41 | 2.24 | 5.65 | 2.21 |
| IQ | 91.4 | 19.3 | 79.48 | 22.9 | 88.03 | 21.01 |
| Educational setting | | | | | | |
| Special education | 93 | 51.6 | 24 | 38.7 | 117 | 48.3 |
| Mainstream | 82 | 45.5 | 37 | 59.6 | 119 | 49.2 |
| Home instruction | 5 | 2.7 | 1 | 1.6 | 6 | 2.5 |
| Sex | | | | | | |
| Male | 158 | 87.8 | 53 | 85.5 | 211 | 87.19 |
| Female | 22 | 12.2 | 9 | 14.5 | 31 | 12.81 |
| Diagnosis | | | | | | |
| Autistic disorder | 125 | 67.8 | 51 | 82.3 | 176 | 72.72 |
| PDD-NOS | 50 | 27.8 | 9 | 14.5 | 59 | 24.38 |
| Asperger's disorder | 5 | 2.8 | 2 | 3.2 | 7 | 2.89 |
| CGI-S | | | | | | |
| 4 (moderately ill) | 64 | 35.6 | 24 | 38.7 | 88 | 36.36 |
| 5 (markedly ill) | 90 | 50.0 | 34 | 54.8 | 124 | 51.24 |
| 6 (severely ill) | 26 | 14.4 | 4 | 6.5 | 30 | 12.40 |
| Race | | | | | | |
| White | 156 | 86.7 | 40 | 64.5 | 196 | 80.99 |
| African American | 15 | 8.3 | 11 | 17.7 | 26 | 10.74 |
| Asian | 6 | 3.3 | 5 | 8.1 | 11 | 4.55 |
| Pacific Islander | 2 | 1.1 | 2 | 3.2 | 4 | 1.65 |
| Other/mixed | 1 | 0.6 | 4 | 6.4 | 5 | 2.07 |
| IQ level | | | | | | |
| $\leqslant$70 | 46 | 13.33 | 21 | 33.9 | 45 | 18.60 |
| $>$70 | 134 | 70.56 | 39 | 62.9 | 166 | 68.60 |

RUBI: Research Units on Behavioral Intervention; SD: standard deviation; PDD-NOS: pervasive developmental disorder-not otherwise specified; CGI-S: Clinical Global Impressions Severity.

measured using Cronbach's $\alpha$, for the ABC subscales in this sample was as follows: Irritability ($\alpha$=0.84), Social Withdrawal ($\alpha$=0.89), Stereotypic Behavior ($\alpha$=0.87), Hyperactivity/Non-compliance ($\alpha$=0.88), and Inappropriate Speech ($\alpha$=0.72).

*CY-BOCS-PDD.* The CY-BOCS-PDD is a modified version of the CY-BOCS developed for use in children with obsessive-compulsive disorder (Scahill et al., 1997). The modified version is a semi-structured clinician-rated scale designed to rate the current severity of repetitive behavior in children with ASD. Once the current repetitive behaviors are identified, they are rated on five items: Time Spent, Interference, Distress, Resistance, and Control. Each of these items is scored on a 5-point scale from 0 (least symptomatic) to 4 (most symptomatic), yielding a total score from 0 to 20. The CY-BOCS has established reliability and validity (Scahill et al., 2006). Internal consistency (Cronbach's $\alpha$) for the CY-BOCS in the current sample was 0.85.

*Vineland Adaptive Behavior Scales II – Parent/Caregiver Rating Form.* The Vineland Adaptive Behavior Scales (VABS) assess adaptive functioning across three domains, Socialization, Communication, and Daily Living Skills, and they include an Adaptive Behavior Composite. The VABS Parent/Caregiver form relies on the parent to rate what the child *does* (as opposed to what the child is capable of) in the course of daily living. The scales have been standardized (mean of 100 $\pm$ 15). Children with ASD consistently have Vineland scores that are one standard deviation or more below their IQs (Paul et al., 2004; Williams et al., 2006). This gap between IQ and adaptive functioning often widens in children with ASD who also have disruptive behavior (Scahill et al., 2012; Williams et al., 2006).

*Early Childhood Inventory-4/Child and Adolescent Symptom Inventory-4.* The Early Childhood Inventory-4 (ECI-4) and Child and Adolescent Symptom Inventory-4 (CASI-4) (Gadow and Sprafkin, 2000, 2005) are behavior rating scales based on the classification and symptoms of the DSM-IV-TR (APA, 2000). Whereas the ECI-4 is used for preschoolers (ages 3–5 years), the CASI-4 is intended for children of primary-school and older, through adolescence. Item content was the same for the subscales used in this

**Table 2.** Factor loadings, test–retest ICC reliability and mean severity ratings for HSQ-ASD items.

| HSQ-ASD item no. and content | Factor 1 (Socially Inflexible | Factor 2 (Demand Specific) | ICC test–retest reliability[a] | Mean severity ratings |
|---|---|---|---|---|
| 1. While playing with other children | **0.47** | −0.08 | 0.429** | 3.72 |
| 9. While you are on the telephone | **0.43** | 0.16 | 0.772** | 4.34 |
| 10. When visitors are in your home | **0.74** | −0.08 | 0.720** | 3.38 |
| 11. When you are visiting someone's home | **0.80** | −0.11 | 0.541** | 4.26 |
| 12. In public places | **0.66** | −0.02 | 0.517** | 5.38 |
| 13. When father (other caregiver) is home | **0.40** | −0.04 | 0.479** | 3.10 |
| 17. When with a babysitter | **0.44** | 0.12 | 0.404* | 2.41 |
| 19. Response to household rules | **0.41** | 0.29 | 0.552** | 4.71 |
| 22. When asked to move from one activity to another | **0.46** | 0.25 | 0.475* | 5.15 |
| 23. When there is an unexpected change in daily routine | **0.45** | 0.18 | 0.834** | 4.94 |
| 25. When attending a large group event | **0.59** | 0.12 | 0.572** | 5.08 |
| 27. When taken to necessary appointment | **0.40** | 0.22 | 0.579** | 3.53 |
| Test–retest reliability of subscale | | | 0.573** | |
| 2. *When asked to put away toys, books, or other personal items* | 0.29 | **0.42** | 0.313* | 4.57 |
| 3. *When asked to come to dinner table* | 0.05 | **0.53** | 0.556** | 3.68 |
| 5. Getting dressed | −0.07 | **0.73** | 0.595** | 3.48 |
| 6. Washing and bathing | 0.01 | **0.64** | 0.337* | 3.03 |
| 7. *When needing to use the toilet* | 0.06 | **0.42** | 0.822** | 3.09 |
| 8. *When told to brush teeth* | −0.08 | **0.52** | 0.709** | 3.10 |
| 15. *When asked to wash hands at meal and other times* | −0.08 | **0.64** | 0.457* | 2.45 |
| 16. At bedtime | 0.11 | **0.59** | 0.650** | 3.57 |
| 18. Getting up in the morning | −0.02 | **0.61** | 0.624** | 2.03 |
| 20. Getting ready to go to school | −0.02 | **0.70** | 0.512** | 3.63 |
| 21. *When asked to put clothes (including jackets, shoes, etc) in proper places* | 0.03 | **0.62** | 0.644** | 3.58 |
| 24. *When asked to get ready to leave the house* | 0.28 | **0.42** | 0.691** | 3.47 |
| Test–retest reliability of subscale | | | 0.575** | |
| 14. When asked to do chores | 0.34 | 0.32 | 0.599** | 4.22 |
| 26. When repetitive behavior is interrupted | 0.27 | 0.34 | 0.364 | 4.21 |
| 4. At meal times | 0.27 | 0.22 | 0.606** | 4.03 |

ICC: Intraclass correlation coefficient; HSQ-ASD: Home Situations Questionnaire-Autism Spectrum Disorder.
Factor loading cut-off criterion was ⩾0.35; seven newly added items are italicized. Factor loadings in bold print are those contributing to each factor (i.e. loading ⩾0.35).
[a]$n = 29$ for test–retest reliability assessment.
*$p \leqslant 0.05$; **$p \leqslant 0.01$.

study, although wording was slightly different to conform to different age levels. For both scales, parents rated symptoms on a 4-point Likert Scale ranging from 0 (never) to 3 (very often). Items measuring inattentive and hyperactive-impulsive symptoms of ADHD (18 items), Oppositional Defiant Disorder (ODD; 8 items), and ASD (12 items) were used in this study. These items represent summaries of the symptoms found in the current edition of the DSM (APA, 2013). The validity of the ECI-4 has been shown in preschool children with ASD (Lecavalier et al., 2011). For the current sample, internal consistency measured using Cronbach's $\alpha$ was as follows for the ECI-4 subscales: ADHD ($\alpha = 0.86$), ODD ($\alpha = 0.87$), and PDD ($\alpha = 0.79$). Cronbach's $\alpha$ for the CASI-4 subscales were as follows: ADHD ($\alpha = 0.78$), ODD ($\alpha = 0.89$), and PDD ($\alpha = 0.74$).

*IQ; Stanford–Binet Intelligence Scales—Fifth Edition.* The Stanford–Binet Intelligence Scales—Fifth Edition (SB-5; Roid, 2003) assesses intelligence and cognitive strengths and weaknesses of individuals aged 2–85 years. Its 10 subtests assess five factors of cognitive ability: Fluid Reasoning, Knowledge, Quantitative, Visual–Spatial, and Working Memory. The SB is a well-established IQ test that has been used in previous RUPP studies. Some children with low IQs were assessed with the Mullen Scales of Early Learning (Mullen, 1995). For the RUBI sample, IQ scores were available for 163 children (151 assessed using abbreviated

IQ on SB and 12 assessed using Mullen Scales). For the Guanfacine sample, IQ scores were available for 62 children (60 assessed using SB and two assessed using Mullen). Since the vast majority of children in this study were assessed on the SB; for consistency, we only analyzed SB scores in our comparisons of divergent validity (see below).

### Procedure

In both the RUBI and Guanfacine studies, IQ was assessed at screening visits, and all other measures were collected at baseline, which usually occurred 1 or 2 weeks later. In order to obtain test–retest reliability for the HSQ-ASD, we asked parents of the last 29 RUBI participants to complete the instrument at the screening visit and again at baseline. The average time interval between ratings was 17.6 days (SD = 9.7; range 6–43).

### Statistical procedures

EFA, using the Comprehensive Exploratory Factor Analysis Software (Browne et al., 2002), was performed to determine the factor structure of the expanded 27-item HSQ-ASD. The polychoric correlation matrix for HSQ item severity scores (range 0–9) was submitted for EFA. Ordinary Least Squares (OLS) was used as the discrepancy function. This is preferred over Maximum Likelihood or Generalized Least Squares Estimation for non-normal data (Norris and Lecavalier, 2010), which was the case with HSQ data in the current sample. Oblique Crawford–Ferguson quartimax rotations were used, as we predicted that latent factors related to non-compliance would be correlated. There is no fixed rule for selecting cut-off factor loadings. We chose a factor loading of $\geqslant 0.35$ as the cut-off for retaining items because it separated clinically meaningful item groupings from items that appeared unrelated. Next, using the Statistical Package for the Social Sciences, Version 19 (SPSS, 2010), we calculated (a) Cronbach's $\alpha$ to assess internal consistency of obtained subscales, (b) intraclass correlations (ICCs) to measure test–retest reliability, and (c) Pearson product-moment correlations between HSQ-ASD subscales and other criterion measures (ABC, VABS, Child and Adolescent Symptom Inventory (CASI)/Early Childhood Symptom Inventory (ECSI), CY-BOCS) and SB IQ to analyze convergent and divergent validity.

## Results

### EFA

Examination of the scree plot (Catell, 1966) supported the choice of a two-factor model in this sample, with an "elbow" at the second data point. The Root Mean Square Error of Approximation (RMSEA)—a measure of model fit indicating the degree of discrepancy that would be found if the model were fit to the population from which the sample was drawn—for the two-factor solution was estimated at 0.06 (90% confidence interval 0.049–0.072). This indicates a "reasonable" fit, per guidelines proposed by Browne and Cudeck (1993). These two factors correspond, with minor differences, to the previously reported (Socially Inflexible and Demand Specific) factors.

Rotated factor loadings of the 27 items appear in Table 2. Factor loadings in bold print are those contributing to each factor (i.e. loading $\geqslant 0.35$). Three items (no. 14, "When asked to do chores," no. 26, "When repetitive behavior is interrupted," and no. 4, "At meal times") did not meet the cut-off of 0.35, even though they were endorsed at rates similar to the other 24 items (79.2%, 75.8%, and 80.8%, compared to 45.0%–92.9%, with a mean of 73.4%, for all items). In addition, these items loaded similarly on both factors, suggesting a lack of specificity. Consequently, these items were discarded from the final composition of the HSQ-ASD. In Chowdhury et al. (2010), item no. 14 previously loaded with the Demand-Specific factor, whereas items no. 4 and 26 previously loaded with the Socially Inflexible factor. Hence, 17 of 20 items (85%) that were in the original HSQ-PDD were confirmed as being on the same respective factors in this analysis. In addition, all seven new items (i.e. the italicized items in Table 2; 100%) that were crafted to assess response to demands did, in fact, load on the Demand-Specific factor.

Based on the EFA, we proceeded with psychometric examination of the 24-item HSQ items, with 12 items on each subscale. The mean factor loadings on the Socially Inflexible and Demand-Specific subscales were 0.52 and 0.57, respectively. As expected, the two subscales were moderately correlated ($r = 0.51$), indicating that these latent factors of non-compliance sampled some aspects of the same overarching domain while also assessing unique attributes of non-compliance. Internal consistency, measured using Cronbach's $\alpha$, was 0.84 for Socially Inflexible and 0.89 for Demand Specific.

Test–retest reliability analysis ($n = 29$) using ICC correlations on HSQ severity ratings showed significant correlations between ratings made at Screen and Baseline for all but one items (no. 26; (Table 2, fourth column)). This item ("When repetitive behavior is interrupted") was discarded from the scale based on the EFA, as described earlier. Test–retest reliability for the subscale totals were all significant with $r = 0.57$ for Socially Inflexible, $r = 0.58$ for Demand Specific, and $r = 0.57$ for the combined total. Test–retest reliability was also assessed using Pearson correlations and yielded similar values as ICC correlations. Finally, subscale and Total scores from the 20-item HSQ-PDD version (Chowdhury et al., 2010) and the current 24-item HSQ-ASD were strongly correlated ($r = 0.98$ for Socially Inflexible subscale, $r = 0.92$ for Demand-Specific subscale, $r = 0.97$ for HSQ Total, all with $p < 0.01$).

In order to provide some reference values for our sample of children with disruptive behaviors, we computed

**Table 3.** Descriptive statistics for HSQ-ASD scores stratified by RCT sample, gender, and intellectual level.

| | Socially Inflexible subscale | | | Demand-Specific subscale | | | HSQ-ASD total | | |
|---|---|---|---|---|---|---|---|---|---|
| | *M* | SD | Range | *M* | SD | Range | *M* | SD | Range |
| RCT sample | | | | | | | | | |
| RUBI ($n=180$) | 51.98 | 20.19 | (2, 97) | 40.77 | 20.63 | (2, 97) | 92.77 | 36.77 | (5, 190) |
| Guanfacine ($n=62$) | 43.96 | 19.99 | (6, 90) | 36.30 | 21.50 | (3, 105) | 80.27 | 35.17 | (13, 154) |
| Gender | | | | | | | | | |
| Male ($n=210$) | 48.98 | 20.27 | (2, 97) | 38.77 | 20.32 | (2, 105) | 87.76 | 36.04 | (5, 180) |
| Female ($n=30$) | 56.96 | 20.25 | (14, 93) | 45.86 | 23.96 | (3, 97) | 102.83 | 39.25 | (40, 190) |
| IQ level | | | | | | | | | |
| $\leqslant 70$ ($n=43$) | 48.74 | 20.49 | (13, 90) | 39.41 | 22.41 | (4, 85) | 88.16 | 39.36 | (17, 154) |
| $> 70$ ($n=166$) | 49.96 | 21.01 | (2, 97) | 40.22 | 20.71 | (2, 105) | 90.19 | 37.04 | (5, 190) |

RCT: randomized clinical trial; HSQ-ASD: Home Situations Questionnaire-Autism Spectrum Disorder; SD: standard deviation; RUBI: Research Units on Behavioral Intervention.

mean severity ratings for each item, and these figures appear in the right column of Table 2. The mean severity score for the Socially Inflexible subscale was 49.98 (SD = 20.40), for the Demand-Specific subscale was 39.65 (SD = 20.89), and HSQ Total score was 89.64 (SD = 36.71). Additional information about HSQ-ASD subscale and Total scores stratified by RCT sample (RUBI vs Guanfacine), gender, and IQ level appear on Table 3. Average HSQ scores in these categories should be interpreted with caution as our sample was not meant to be normative. In particular, the number of participants in some of the subcategories were too small to allow for meaningful interpretation (e.g. only 13% of participants were female) and overall were not representative.

All validity analyses were conducted using the new, 12-item HSQ-ASD subscale scores and the new 24-item HSQ-ASD Total score. To evaluate convergent validity, correlations were calculated between HSQ scores and scores on the following criterion measures: (a) parent-rated ABC; (b) clinician-rated CY-BOCS Total Score; (c) ADHD, ODD, and PDD subscales of the parent-rated CASI and ECSI; and (d) Vineland Daily Living domain. Divergent validity was assessed by examining correlations with IQ scores and Vineland Communication and Socialization domains (see Table 4). As expected, HSQ subscale and Total score had significant correlations with the measures of problem behavior (four of the five ABC subscales, CY-BOCS Total Score, and CASI/ECSI subscales). The Socially Inflexible scores were more strongly correlated than Demand Specific with the ABC Irritability subscale and the PDD subscale of the CASI/ECSI. On the other hand, Demand-Specific scores had stronger correlations with the VABS Daily Living subscale. Neither subscale was correlated with SB IQ score or the Vineland Communication and Socialization domains. As far as correlations between age and HSQ subscale and total scores, only the Socially Inflexible subscale had a statistically significant association, this being a low inverse relation with age ($r=-0.16$, $p<0.05$).

## Discussion

The two subscales, Socially Inflexible and Demand Specific, derived in this psychometric study of the HSQ-ASD should provide better interpretation of scores in the ASD population, thus increasing the clinical utility of the instrument. The underlying themes of these two subscales suggest that Social Inflexibility is congruent with deficits in social interactions and rigid adherence to routines that are part of ASD, whereas Demand-Specific non-compliance appears to reflect oppositional behavior in response to typical daily requests. The content of these two subscales is clinically and theoretically meaningful in the context of situational non-compliance in this population. Although children in both the RUBI and the Guanfacine studies were seeking treatment for different disruptive behaviors, combining the samples broadened the characteristics of the resulting study sample. For example, the age of children in the RUBI study (mean = 4.7 years) was younger than ages in the Guanfacine study (mean = 8.4 years), with the result that the combined sample was representative of a broader age range. The two-factor structure derived in this study was similar to that obtained in our psychometric examination of the previous 20-item HSQ-PDD (Chowdhury et al., 2010), and 19 of the 20 items from the HSQ-PDD loaded onto the same factors as they did in our previous study, with some minor differences in the loadings of some items. Overall, findings from the EFA suggested good construct validity of the HSQ-ASD.

We mostly succeeded in our attempt to increase the number of items in the Demand-Specific subscale, which contained only six items based on our previous factor analytic investigation (Chowdhury et al., 2010). The EFA confirmed that 6 of our 7 new items loaded on the Demand-Specific factor. One pre-existing "demand" item (no. 16: "When asked to do chores") failed to reach the specified factor loadings for retention. The resulting 24-item HSQ-ASD has an equal number of items in the two subscales giving equal opportunity for each subscale to contribute to

**Table 4.** Pearson correlation coefficient between HSQ-ASD subscales and total score and criterion measures at baseline.

| Criterion | Socially Inflexible subscale | Demand-Specific subscale | HSQ-ASD total |
|---|---|---|---|
| ABC subscales (parent-rated) ($n=240$) | | | |
| Irritability | 0.506** | 0.321** | 0.464** |
| Social Withdrawal | 0.303** | 0.283** | 0.329** |
| Stereotypic Behavior | 0.188** | 0.206** | 0.222** |
| Hyperactivity/Non-compliance | 0.382** | 0.368** | 0.422** |
| Inappropriate Speech | 0.053 | 0.122 | 0.099 |
| Vineland domains | | | |
| Communication ($n=239$) | −0.091 | −0.105 | −0.110 |
| Daily Living ($n=230$) | −0.188** | −0.258** | −0.251** |
| Socialization ($n=235$) | −0.107 | −0.061 | −0.094 |
| Adaptive Behavior Composite ($n=225$) | −0. 049 | −0.012 | −0.034 |
| CASI and ECSI subscales ($n=240$) | | | |
| ADHD | 0.167* | 0.176** | 0.193** |
| ODD | 0.266** | 0.228** | 0.278** |
| PDD | 0.273** | 0.160* | 0.243** |
| CY-BOCS Total Score ($n=240$) | 0.258** | 0.263** | 0.293** |
| IQ ($n=223$) | −0.016 | 0.041 | 0.014 |

HSQ-ASD: Home Situations Questionnaire-Autism Spectrum Disorder; ABC: Aberrant Behavior Checklist; CASI: Child and Adolescent Symptom Inventory; ECSI: Early Childhood Symptom Inventory; CY-BOCS: Children's Yale–Brown Obsessive-Compulsive Scale; ODD: oppositional defiant disorder; ADHD: attention deficit hyperactivity disorder; PDD: pervasive developmental disorder.
*$p \leq 0.05$; **$p \leq 0.01$.

the total score. Internal consistencies were satisfactory for both subscales and comparable to levels reported by Chowdhury et al. (2010). Test–retest reliability, computed using a small subset of $n=29$ participants, was statistically significant and adequate for individual items. The figures for test–retest reliability for subscale scores and total score were modest (albeit statistically significant). This may be due in part to attenuation to the range of scores, as children who did not have disruptive behavior were not admitted to the RUBI study, essentially precluding low HSQ scores from our sample. The item mean severity score for HSQ Total ($3.73 \pm 1.53$) and subscale scores ($4.16 \pm 1.70$ and $3.30 \pm 1.74$) for Socially Inflexible scales and Demand Specific, respectively, provides a reference point for scores for children with ASD seeking treatment for disruptive behavior problems.

Validity analyses using the per-item mean HSQ-ASD score and the two subscale scores provided evidence of convergent and divergent validity. As predicted, both HSQ total and subscale scores had significant positive correlations with ABC Irritability and Hyperactivity suggesting convergent validity. Additional evidence of convergent validity was provided by significant positive correlations between the HSQ subscale scores and total, CASI/ECSI subscales of ADHD, ODD, PDD, and CY-BOCS total. Correlations between HSQ-ASD scores and VABS Daily Living Skills were significant, but modest. As expected, correlations between HSQ scores and IQ scores plus Vineland Communication and Socialization domains were non-significant, providing evidence of divergent validity.

In summary, the modified and expanded 24-item HSQ-ASD provides broader coverage of situations associated with non-compliance in young children with ASDs. This should increase the clinical utility of the instrument. In addition, the two derived subscales should help to identify situational differences in non-compliance and in turn facilitate a more refined interpretation of non-compliant behaviors. For instance, high Socially Inflexible scores in the presence of relatively low Demand-Specific scores may reflect a predominance of rigid adherence to routines and need for sameness that are influencing non-compliant behavior.

Limitations of this study include the following. First, all participants were selected for presence of behavior problems (particularly irritability, tantrums, aggression, and hyperactivity), which limits generalizability of the findings. Although the combined sample broadens the representativeness of the sample beyond either sample alone, the findings may only generalize to children with ASD (ages 3–14 years) accompanied by disruptive behavior. Similarly, the mean HSQ total and subscale scores that we provided as reference points are clearly much higher than would be seen in a random sample of children with ASD.

As noted in Chowdhury et al. (2010), there are few instruments for assessing non-compliance in children with ASD. Given that non-compliance is a frequently reported behavioral problem in ASD (Lecavalier, 2006), development of a reliable and valid outcome measure is overdue. This study provides psychometric evidence that the HSQ-ASD performs well for assessing non-compliance and the

situational/demand triggers of non-compliance in preschool and school-age children with ASD. Future studies will provide a valuable service by evaluating the HSQ-ASD in children with ASD unselected for disruptive behavior. It would also be useful to have psychometric data from adolescents on the HSQ-ASD as done by Adams et al. (1995), who designed adolescent parent-report and self-report forms of the HSQ for typically developing children (AHSQ-pr and AHSQ-sr, respectively). Findings from this study support the use HSQ-ASD as an outcome measure for disruptive behaviors in this population.

## Funding

The RUBI study was funded by the National Institute of Mental Health grants to Yale University/Emory University (MH081148; principal investigator: Dr Scahill), the University of Pittsburgh (MH080965; principal investigator: Dr Johnson), Ohio State University (MH081105; principal investigator: Dr Lecavalier), Indiana University (MH081221; principal investigator: Dr Swiezy), and the University of Rochester (MH080906; principal investigator: Dr Smith). The Guanfacine study was funded by National Institute of Mental Health grants: to Dr Scahill (R01MH083707), Dr McDougle (RO1MH83739), Dr McCracken (RO1MH083747), and Dr King (R01MH86927).

## References

Adams CD, McCarthy M and Kelly ML (1995) Adolescent versions of the Home and School Situations Questionnaires: initial psychometric properties. *Journal of Clinical Child Psychology* 24: 377–385.

Aman MG, McDougle CJ, Scahill L, et al.; Research Units on Pediatric Psychopharmacology (2009) Medication and parent training in children with pervasive developmental disorders and serious behavior problems: results from a randomized clinical trial. *Journal of the American Academy of Child and Adolescent Psychiatry* 48: 1143–1154.

Aman MG, Singh NN, Stewart AW, et al. (1985a) Psychometric characteristics of the Aberrant Behavior Checklist. *American Journal of Mental Deficiency* 89: 492–502.

Aman MG, Singh NN, Stewart AW, et al. (1985b) The Aberrant Behavior Checklist: a behavior rating scale for the assessment of treatment effects. *American Journal of Mental Deficiency* 89: 485–491.

American Psychiatric Association (APA) (2000) *Diagnostic and Statistical Manual IV (DSM IV)*. Washington, DC: APA.

American Psychiatric Association (APA) (2013) *Diagnostic and Statistical Manual-5 (DSM 5)*. Washington, DC: APA.

Arnold LE, Aman MG, Li X, et al. (2012) Research units of pediatric psychopharmacology (RUPP) autism network randomized clinical trial of parent training and medication: One-year follow-up. *Journal of the American Academy of Child & Adolescent Psychiatry* 51(11): 1173–1184.

Arnold LE, Vitiello B, McDougle C, et al. (2003) Parent-defined target symptoms respond to risperidone in RUPP autism study: customer approach to clinical trials. *Journal of the American Academy of Child and Adolescent Psychiatry* 42: 1443–1450.

Barkley RA and Edelbrock C (1987) Assessing situational variation in children's problem behaviors: the Home and School Situations Questionnaires. In: Prinz R (ed.) *Advances in Behavioral Assessment of Children and Families*. Greenwich, CT: JAI Press Inc., pp. 157–176.

Bearss K, Johnson C, Smith T, et al. (2015) Parent training for young children with autism spectrum disorder and disruptive behavior: A randomized clinical trial. *Journal of the American Medical Association* 313(15): 1524–1533.

Breen MJ and Altepeter S (1991) Factor structures of the Home Situations Questionnaire and the School Situations Questionnaire. *Journal of Pediatric Psychology* 16: 59–67.

Browne MW and Cudeck R (1993) Alternative ways of assessing model fit. In: Bollen KA and Scott Lang J (eds) *Testing Structural Models*. Newbury Park, CA: SAGE, pp. 136–162.

Browne MW, Cudeck R, Tateneni K, et al. (2002) CEFA: comprehensive exploratory factor analysis, Version 3.01 (Computer software and manual). Available at: http://faculty.psy.ohio-state.edu/browne/programs.htm (accessed 23 November 2008).

Cattell RB (1966) The scree test for the number of factors. *Multivariate Behavioral Research* 1: 245–276.

Chowdhury M, Aman MG, Scahill L, et al. (2010) The Home Situations Questionnaire–PDD version: factor structure and psychometric properties. *Journal of Intellectual Disabilities Research* 54: 281–291.

DuPaul GJ and Barkley A (1992) Situational variability of attention problems: psychometric properties of the revised Home and School Situations Questionnaires. *Journal of Clinical Child Psychology* 21: 178–188.

Gadow KD and Sprafkin J (2000) *Early Childhood Symptom Inventory-4 Screening Manual*. Stony Brook, NY: Checkmate Plus.

Gadow KD and Sprafkin J (2005) *Child and Adolescent Symptom Inventory-4R*. Stony Brook, NY: Checkmate Plus.

Guy W (ed.) (1976) *ECDEU Assessment Manual for Psychopharmacology*. Rockville, MD: US Department of Heath, Education and Welfare Public Health Service Alcohol, Drug Abuse and Mental Health Administration.

Kaat AJ, Lecavalier L and Aman MG (2014) Validity of the aberrant behavior checklist in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 44(5): 1103–1116.

Lecavalier L (2006) Behavioral and emotional problems in young people with pervasive developmental disorders: relative prevalence, effects of subject characteristics, and empirical classification. *Journal of Autism and Developmental Disorders* 36: 1101–1114.

Lecavalier L, Gadow KD, DeVincent CJ, et al. (2011) Validity of DSM-IV syndromes in preschoolers with autism spectrum disorders. *Autism* 15: 527–543.

Lord C, Risi S, Lambrecht L, et al. (2000) The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism & Developmental Disorders* 30: 205–223.

Mullen EJ (1995) *Mullen Scales of Early Learning*. Bloomington, MN: Pearson Assessments.

Norris M and Lecavalier L (2010) Evaluating the use of exploratory factor analysis in developmental disability psychological research. *Journal of Autism and Developmental Disorders* 40: 8–20.

Paul R, Miles S, Cicchetti D, et al. (2004) Adaptive behavior in autism and pervasive developmental disorder-not otherwise specified: microanalysis of scores on the Vineland Adaptive Behavior Scales. *Journal of Autism & Developmental Disorders* 34: 223–228.

Roid GH (2003) *Stanford-Binet Intelligence Scales*. 5th ed. Itasca, IL: Riverside Publishing.

Scahill L, Aman MG, McDougle CJ, et al. (2009) Trial design challenges when combining medication and parent training in children with pervasive developmental disorders. *Journal of Autism and Developmental Disorders* 39: 720–729.

Scahill L, MacDougle CJ, Aman MG, et al.; Research Units on Pediatric Psychopharmacology Autism Network (2012) Effects of risperidone and parent training on adaptive functioning in children with a pervasive developmental disorders and serious behavioral problems. *Journal of the American Academy of Child and Adolescent Psychiatry* 51(2): 136–146.

Scahill L, MacDougle CJ, Williams S, et al. (2006) The Children's Yale-Brown Obsessive Compulsive Scales modified for pervasive developmental disorders. *Journal of the American Academy of Child and Adolescent Psychiatry* 45: 1114–1123.

Scahill L, Riddle MA, McSwiggin-Hardin M, et al. (1997) Children's Yale-Brown Obsessive Compulsive Scale: reliability and validity. *Journal of the American Academy of Child and Adolescent Psychiatry* 36: 844–852.

SPSS (2010) *Statistical Package for the Social Sciences* (Version 19.0, computer software). Chicago, IL: SPSS Inc.

Williams SK, Scahill L, Vitiello B, et al. (2006) Risperidone and adaptive behavior in children with autism. *Journal of the American Academy of Child and Adolescent Psychiatry* 45(4): 431–439.