Factors influencing the severity of behavioral phenotype in autism spectrum disorders: Implications for research

Mingyang Zou[a],[1], Caihong Sun[a],[1], Jia Wang[a], Jing Kang[a], Zhanbin Xu[a], Yongjuan Ma[a], Lei Chen[b], Xinxi Zhang[c], Wei Xia[a],[*], Lijie Wu[a],[*]

[a] *Department of Children's and Adolescent Health, Public Health College, Harbin Medical University, Harbin 150081, China*
[b] *Department of Child Health Care, Dalian Maternal and Child Health Care Hospital, Dalian 116033, China*
[c] *Department of Child Health Care, Yangzhou Maternal and Child Health Care Hospital, Yangzhou 225002, China*

ARTICLE INFO

ABSTRACT

The phenotypic heterogeneity of Autism Spectrum Disorders (ASD) presents particular research challenges in the assessment of symptom severity, while the standardized Autism Diagnostic Observation Schedule (ADOS) scores present a severity metric, namely calibrated severity scores (CSS) that are relatively impervious to individual characteristics. To date, no studies have examined the convergent validity of CSS in Chinese sample populations. The present study investigated the validity of the ADOS-CSS using a sample of 321 children aged 2–18 years with ASD, and developed upon existing literature examining the influence of non-ASD-specific characteristics on other types of measures including Autism Diagnostic Interview-Revised (ADI-R), Social Responsiveness Scale (SRS), and Vineland Adaptive Behavior Scales (VABS). As expected, the findings revealed that the CSS were less influenced than ADOS-RAW scores by the demographic and developmental-level variables. Moreover, compared to the ADOS-CSS, the ADI-R, SRS and VABS were still strongly correlated with confounding factors, such as chronological age, intelligence quotients, and language-level. The results of this study corroborate the utilization of CSS as a more valid indicator of ASD severity than raw scores from ADOS and other instruments.

1. Introduction

Autism spectrum disorders (ASD) are behaviorally defined neurodevelopmental disabilities characterized by impairments in reciprocal social interaction and communication as well as restricted, repetitive patterns of behavior or interests (American Psychiatric Association, 2013). Measuring the relative severity of ASD-specific features may assist in developing an accurate description of ASD phenotypes across sample populations from different studies, tracking the progress of the disorder over time, and quantifying treatment outcomes (Havdahl et al., 2016). However, the extreme heterogeneity in phenotypic presentation of ASD poses investigative and clinical challenges for the assessment of symptom severity. This heterogeneity may be attributable at least in part to differences in individual characteristics unrelated to ASD core features, such as chronological age (CA), intelligence quotients (IQ), and language-level, which affect the behavioral phenotypes of ASD, potentially masking the true severity of ASD symptoms.

Numerous studies examining ASD have relied directly on scores from diagnostic and screening measures of the behavioral phenotypes

in order to link biomarkers such as genetic abnormalities, patterns of neural activity or current brain function, or to measure treatment outcomes (Ousley et al., 2017; Tu et al., 2016). The current gold standard for ASD diagnosis is the use of the Autism Diagnostic Interview-Revised (ADI-R) and the Autism Diagnostic Observation Schedule (ADOS). While higher ADI-R and ADOS raw scores seem reliable in suggesting a greater degree of impairment, the scores vary in the extent to which they are in association with both IQ and the CA of the child at the time the interview was conducted (de Bildt et al., 2004). Furthermore, the scores were neither intended nor appropriate for interpretation as a severity metric. One particular restriction of the ADI-R scores as a measure of ASD severity is that scores related to approximately 25% of the total number of items are not obtained for nonverbal children. Thus, the scores of communication domain are limited by nonrandom missing value (Gotham et al., 2009; Hus and Lord, 2013). Previous research has demonstrated that the ADI-R total or domain scores are affected by individual characteristics (de Bildt et al., 2015; Hus and Lord, 2013). Nonetheless, these scores are often used, without modification, as indicators of overall or domain-specific ASD severity

([Ousley et al., 2017](#)). The ADOS represents an observation of children's behaviors in answer to various social stimuli, with the four modules tailored to the children's language-level and CA. Although these modules were created for purpose of controlling for the predominant influence of expressive language-level on behaviors, differences in the constituent and number of items within the algorithms across the modules have restricted the degree to which scores can be compared across modules over time as well as across individuals. In this regard, several common measures present the ASD-related severity ratings such as the Childhood Autism Rating Scale, the Autism Behavior Checklist, and Social Responsiveness Scale (SRS). However, such measures are liable to generate scores that are not independent of individual characteristics. The scores yielded by the Childhood Autism Rating Scale and the Autism Behavior Checklist are either notably in association with IQ or inconsistent with standard measures of diagnosis ([Szatmari et al., 2003](#)). Although a higher SRS score indicates more severe social deficits and autistic behaviors, research has indicated that higher SRS total scores were also observed in older children as well as children with greater expressive language impairment and cognitive delays ([Channell et al., 2015; Hus et al., 2013](#)). When viewed collectively, the scores from these instruments are problematic indicators of ASD symptom severity, and insufficient consideration of potentially confounding factors can pose inaccurate or misleading consequences for the interpretation of findings, including an erroneous estimation of the association between ASD and etiological factors. Consequently, there is a substantial need to continue to refine and improve upon the performance of available measures.

Recently, Gotham et al. ([Gotham et al., 2009, 2007](#)) revised the ADOS algorithm which contains two domains, namely social affect (SA) and restricted and repetitive behaviors (RRB), in order to develop calibrated severity scores (CSS) that more accurately capture core ASD symptom severity. With regard to the same number of items and similar content across modules in Gotham's algorithm, the comparability of 10-point CSS across modules was increased, contributing to use of the ADOS longitudinally to track the severity of symptoms over individuals with different language-levels as well as within individuals when language-level varies according to variables such as age, for example.

To date, although CSS have been widely used by Western research as a severity metric ([de Bildt et al., 2011; Grzadzinski et al., 2016; Pickles et al., 2016; Wiggins et al., 2017](#)), there was a paucity of studies focusing on the utility of CSS in Chinese sample populations. In addition, comparison of ADOS severity with alternative measures of ASD severity is essential to establish the convergent validity of CSS. Moreover, little research has systematically evaluated how non-ASD-specific characteristics influence the scores from other measures of ASD-symptoms. In spite of their implications with regard to the explainability of scores, especially when utilized as indicators of core severity, such potential consequences are rarely acknowledged by researchers using these different types of measures. Therefore, the aim of the present study is to examine the psychometric characteristics of CSS for total and domain scores across ADOS modules 1–3 using an independent sample of Chinese autistic patients. Furthermore, this study develops upon existing literature by evaluating the influences of individual characteristics, containing demographics (gender, maternal education, residence type) and developmental level (CA, IQ and language-level), on other types of measures including the ADI-R, SRS, and Vineland Adaptive Behavior Scales (VABS). This study offers a more thorough understanding of the relationship between potential factors and scores, which is pivotal to determining their utilization as indicators of ASD severity.

2. Materials and methods

2.1. Participants

The sample population consisted of 321 children aged 2–18 years

Table 1
Sample description and scores on primary measures.

Variants	n	Mean ± SD, n (%)	Range
CA (years)	321	7.65 ± 4.14	
Gender (male)	321	262 (82)	
Res type (urban)	321	242 (75)	
Race (Han)	321	306 (95)	
Mat edu (bachelor degree or higher)	321	98 (31)	
Cognitive assessment			
FSIQ	251	53.29 ± 21.99	21–125
VIQ	251	52.44 ± 21.06	19–130
NVIQ	251	59.84 ± 22.74	19–122
Language-level (ADOS Module)			
Module 1	318	200 (63)	
Module 2	318	96 (30)	
Module 3	318	22 (7)	
ADOS			
SA-RAW	318	13.19 ± 3.81	0–20
RRB-RAW	318	3.13 ± 1.73	0–8
ADOS-RAW	318	16.38 ± 4.65	2–26
SA-CSS	318	7.19 ± 1.57	2–10
RRB-CSS	318	6.28 ± 1.97	1–10
ADOS-CSS	318	6.82 ± 1.48	1–10
ADI-R			
ADI-R diagnostic	299	38.46 ± 8.95	12–54
ADI-R current	299	19.96 ± 7.78	1–36
SRS	284	87.30 ± 25.41	17–174
VABS	280	71.50 ± 19.89	25–145
Communication	280	78.86 ± 24.59	45–145
Daily living skills	280	72.21 ± 19.73	22–145
Socialization	280	66.42 ± 17.58	48–145
Motor skills	280	86.15 ± 17.32	54–144

CA = chronological age, Res type = residence type, Mat edu = maternal education, FSIQ = full scale intelligence quotients, VIQ = Verbal IQ, NVIQ = nonverbal IQ, ADOS = Autism Diagnostic Observation Schedule, SA = social affect, RRB = restricted repetitive behavior, CSS = calibrated severity scores, ADI-R = Autism Diagnostic Interview-Revised, SRS = Social Responsiveness Scale, VABS = Vineland Adaptive Behavior Scales.

who were consecutively recruited from the Child Developmental and Behavior Center of Harbin Medical University, Harbin, China. All participants were required to have received an ASD diagnosis from two independent specialist clinicians from our research team using the diagnostic criteria specified in the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-V) ([Association, 2013](#)). Exclusion criteria included (1) neurodevelopmental disorders of known etiology (e.g. Fragile X Syndrome, tuberous sclerosis, or known chromosomal abnormalities or metabolic disorders), (2) significant sensory or motor impairment(e.g. deaf, blindness), and (3) serious chronic diseases.

The sample of participants with ASD comprised those who were 7.65 ± 4.14 years old (male = 262; female = 59) and who were children from rural (25%) and urban areas (75%). Almost all participants were Chinese Han, with the exception of 5% participants from ethnic minorities. 31% of mothers, who represented the primary sources of data for this analysis, were well-educated (i.e. obtained a bachelor degree or higher). The remaining 69% of mothers had attained some college-level or below. Sample characteristics are provided in Table 1.

All participants recruited for the purpose of the research received oral feedback and a written summary of the results of evaluation without any payment. Caregivers provided written informed consent and children provided assent if possible. The procedures related to this study were approved by the Institutional Review Board of Harbin Medical University.

2.2. Measures

In this study, all psychologists or trainees involved in administering the measurement protocols had established research reliability. Each participant was evaluated using the ADI-R, ADOS, SRS, VABS, and a