

## Brief Report: “Quick and (not so) Dirty” Assessment of Change in Autism: Cross-Cultural Reliability of the Developmental Disabilities CGAS and the OSU Autism CGI

Nora Choque Olsson · Sven Bölte

Published online: 31 December 2013  
© Springer Science+Business Media New York 2013

**Abstract** There are few evaluated economic tools to assess change in autism. This study examined the interrater reliability of the Developmental Disabilities Children’s Global Assessment Scale (DD-CGAS), and the OSU Autism Clinical Global Impression (OSU Autism CGI) in a European setting. Using these scales, 16 clinicians with multidisciplinary background and varying experience independently rated eight vignettes of autism spectrum disorder for severity and general psychosocial functioning at referral and discharge. Intraclass correlation coefficient (*ICCs*) for experienced clinicians were .75 for the DD-CGAS and .72 for the OSU Autism CGI. In inexperienced clinicians these *ICCs* were .58 and .59. Results confirm previous North American studies, and further extends the reliability of the instruments to untrained, less experienced clinicians with different professions.

**Keywords** Intervention · Outcome · Longitudinal · Asperger syndrome · Psychometrics

### Introduction

Although research on outcome and treatment response is of paramount importance in child mental health (Arnold et al. 2000), the literature in autism spectrum disorder (ASD) is only slowly improving (Howlin and Moss 2012; Oono et al. 2013). Among the factors hampering better evidence is a lack of the possibility to draw overarching conclusions from different studies, owing to the fact that researchers have not yet adopted a common set of outcome measures (Reichow et al. 2012). Studies have used a diverse repertoire of cognitive tests and clinical scales to determine outcome that mostly were not constructed to assess change in ASD or generally to be particularly sensitive to change (Howlin et al. 2009), for example the Wechsler IQ scales or the Vineland Adaptive Behavior Scales. Often the choice of the instruments has rather reflected the limited availability of more adequate tools for measuring change than its genuine appropriateness.

Although substantial progress in the past two decades concerning the development and psychometric evaluation of screeners, diagnostic rating scales and interviews in ASD has been made (Charman and Gotham 2013), the focus has clearly been on tools for establishing categorical diagnostic status rather than assessing change over time or treatment response. For instance, the “gold standard” of diagnosing ASD, the Autism Diagnostic Observation Schedule-2 (ADOS-2) (Lord et al. 2012; Lord et al. 1994) and Autism Diagnostic Interview-Revised (Rutter et al. 2003) were both developed with the aim to maximize diagnostic/discriminant validity, not to measure change of

---

We sincerely thank the clinicians who participated in this study, as well as Stephanie Hansson for her support in the creation of vignettes. Sven Bölte was supported by the Swedish Research Council (Grant No. 523-2009-7054). Nora Choque Olsson and the current study were supported by Stiftelsen Sunnerdahls Handikappfond, Majblomman, Sällskapet Barnavård, and the Stockholm County Council.

---

N. Choque Olsson · S. Bölte (✉)  
Pediatric Neuropsychiatry Unit, Department of Women’s and Children’s Health, CAP Research Center, Center of Neurodevelopmental Disorders (KIND), Karolinska Institutet, Gävlegatan 22, 11330 Stockholm, Sweden  
e-mail: sven.bolte@ki.se

N. Choque Olsson  
e-mail: nora.choque-olsson@ki.se

N. Choque Olsson · S. Bölte  
Division of Child and Adolescent Psychiatry, Stockholm County Council, Stockholm, Sweden

symptoms or severity over time. Only very recently, there have been attempts to quantify ADOS scores by the creation of a 10-point calibrated severity scale (de Bildt et al. 2011; Gotham et al. 2012) which better enable comparability of score between ADOS modules. In addition, a new version of the ADOS, the ADOS-C (C for change) is in preparation, which hopefully will represent a sensitive metric of symptom or functional capacity assessment over time in the future (Lord et al. 2013). Nevertheless, the ADOS and ADI-R are rather time and cost intensive, and require extensive training and experience. Another widely applied scale in ASD, the Social Responsiveness Scale (Bölte 2012; Bölte et al. 2008; Constantino and Gruber 2012; Constantino and Frazier 2013) is a quick quantitative measure of autistic traits. It provides a standard measure of error for reassessment in order to enable to judge whether changes are random variation or are likely to embody true changes. The SRS has repeatedly demonstrated sensitivity to treatment response, for instance in clinical trials of social skills training (Tse et al. 2007; Wood et al. 2009). However, the SRS is an informant-based measure, not an expert-based measure. Moreover, it is not constructed as a measure of clinical autism severity or functional impairment in everyday life.

In conclusion, there is shortage of economic, intuitive, clinician-based, reliable measures of symptomatic and functional change in ASD. As DSM-5 (APA 2013) introduces severity specifiers for ASD behavior domains, functional impairment and the associated need for support, the availability of such tools will become ever important in the future. In order to facilitate the availability of economic scales to measure change in international ASD research and practice, and to examine their practicability and scientific value, this study aimed at determining the inter-rater reliability and feasibility of the Developmental Disabilities Children's Global Assessment Scale (DD-CGAS) (Wagner et al. 2007) and the OSU Autism Clinical Global Impression (OSU Autism CGI) (RUPP 2005) in a European environment. Both instruments are clinician friendly tools, taking only a few minutes to complete, that have been derived from the well-known and widespread Children Global Assessment Scale, and Clinical Global Impression scale to fit ASD assessment. In addition to previous North American studies, the current article particularly examined the scales' properties in experienced versus inexperienced clinicians with varying professional background, who spontaneously (without explicit training) conducted ratings. The latter might allow to evaluate the scales' scientific properties in more a more conservative and perhaps naturalistic fashion.

## Methods

### Instruments

#### *DD-CGAS*

For the current study, the DD-CGAS was translated into Swedish, back-translated to English and authorized. The DD-CGAS (Wagner et al. 2007) is a clinician-rated scale derived from the Children Global Assessment Scale (Shaffer et al. 1983). Text revisions are introduced to the original CGAS to enable a more targeted functional assessment in ASD. It yields a single score of current global functioning for children and adolescents with ASD aged 4–18 years relative to the typical development of same age peers. Ratings can be completed in less than 5 min. Scores are based on all available sources of information from multiple areas of functioning: self-care, social behavior, and school/academic functioning. Maintaining the overall structure of the original CGAS, the DD-CGAS scores range from 1 (maximum impairment) to 100 (superior level of functioning). The scale is divided into 10-point intervals that are headed by a description of the level of functioning. Each interval (1–10, 11–20 and so forth) has a descriptive header (e. g. “Moderate impairment in functioning in most domains”, for range of score 60–51). Scores below 70 on the DD-CGAS indicate atypical functioning in a clinically relevant range. For administration, the examiner first determines the level of impairment for each domain, taking into account the child's behavior in various environments, as well as the level of environmental accommodation necessary to support the child and the level of support required. Then the rater selects the composite reference range that best corresponds with the level of functioning in all domains (for example, 60–51 moderate operating in most areas). The examples used in the ranges to confirm the description of the child's functioning, although no child will be perfectly described by these descriptions.

Previous North American data on the psychometric properties of DD-CGAS in certified trained raters within ongoing research on pediatric psychopharmacology judging children with pervasive developmental disorders yielded good to excellent interrater reliability [intraclass correlation coefficient ( $ICC$ ) = .79], retest reliability (average  $ICC$  = .86), and concurrent validity with the Aberrant Behavior Checklist-I ( $r$  = .71) and Global Impression Scale-I ( $r$  = .52) (Wagner et al. 2007).

### OSU Autism CGI

The CGI (Guy 1976) is a tool to rate global symptom severity and symptom improvement that takes less than 5 min to complete, and has demonstrated good psychometric properties in large European child and adolescent psychiatric cohorts (Dyrborg et al. 2000; Lundh et al. 2010; Lundh et al. 2012). It combines two ratings, current symptom severity (CGI-S) and symptomatic improvement compared to baseline (CGI-I). Severity (CGI-S) is scored on a 7-point scale from (1) “Normal, not at all ill” to (7) “Among the most extremely ill patients”. The OSU Research Unit on Pediatric Psychopharmacology modified the CGI for individuals with ASD [OSU Research Unit on Pediatric Psychopharmacology (RUPP) 2005] to create the OSU Autism CGI. The scale is rated in a similar way to the classical CGI, but it is focused on autism spectrum symptoms. Symptoms frequently associated with autism spectrum—such as compulsions, hyperactivity, and self-injury should also be considered. As far as we are aware, no psychometric data has yet been published on the OSU Autism CGI. For the current study, the OSU Autism CGI was translated into Swedish, back-translated to English, checked for clarity and piloted. Only the severity scale of the CGI, the CGI-S, was used.

### Raters

To establish the interrater reliability of the Swedish DD-CGAS and OSU Autism CGI-S, 16 clinicians involved in neuropsychiatric assessments on a daily basis from 9 child and adolescent psychiatry outpatient units in Stockholm County with varying professional background, and clinical experience were included: psychologists (11), social workers (3) and nurses (2). There were 13 female and 3 male raters, ranging in age between 31 and 62 years, with a mean age of 42 years. Eleven (69 %) had master and 5 (31 %) bachelor degrees. Their clinical experience ranged from 1 to 40 years, with a mean of 7.8 years. The raters were divided in two groups: *experienced* [ $<2$  years clinical experience to work with ASD = 8 raters (7 psychologists and 1 social worker)] and *inexperienced* [ $>2$  years clinical experience to work with ASD = 8 raters (4 psychologists, 2 nurses, 1 social worker)]. Raters were naïve to both the DD-CGAS and OSU Autism CGI-S, and had neither been explicitly trained on the CGAS or DD-CGAS for this study. They were in the majority not familiar to the original CGI-S, but mostly familiar to the classical CGAS.

### Vignettes

Clinicians rated eight written vignettes of ASD reflecting a range of adaptive functioning and symptom severity using

the DD-CGAS and OSU Autism CGI-S (Table 1). They described 3 girls and 5 boys given a clinical ICD-10 (WHO 1992) consensus diagnoses of autism, Asperger syndrome or pervasive developmental disorder not otherwise specified, who also fulfilled current DSM-5 (APA 2013) and ADOS [module 3 (new algorithm) or module 4)] criteria for ASD. The cases were aged between 8 and 16 years, and had borderline to high range IQs (75–120). Gold standard scores for the DD-CGAS and OSU Autism CGI-S regarding the vignettes were established by averaging the two author’s (Swedish version developers) independent ratings for each of the vignettes. Gold standard ratings of the vignettes ranged from 32 to 69 ( $M = 53.6$ ) for the DD-CGAS and 2–5 ( $M = 3.5$ ) for OSU Autism CGI at referral and 40–75 ( $M = 59.4$ ) for the DD-CGAS and 2–5 ( $M = 3.1$ ) for OSU Autism CGI at discharge.

Each vignette comprised extensive clinical descriptions of the individual’s situation, symptomatology, and treatment for two points in time: one for clinical referral (before treatment), the other for discharge (after treatment). Common interventions were individual cognitive behavior therapy, medication, social skills training, psychoeducation and general information about ASD. Clinicians rated the vignettes separately for both time points resulting in a total of 16 ratings for each of the instruments. Vignettes were based on true clinical cases of ASD at the division of child and adolescent psychiatry, Stockholm County. The vignettes (2–3 pages in length) included information on age, gender, IQ, ASD diagnosis, psychiatric co-morbidities, as well as behavioral descriptions of everyday functioning regarding adaptive skills, social integration, and treatment provided.

### Procedure and Statistics

The 16 raters independently and spontaneously rated the eight clinical vignettes of ASD in terms of psychosocial functioning using the DD-CGAS, and clinical severity using the OSU Autism CGI-S, for referral and discharge. The only instruction given prior to ratings was to read the general information provided on the instruments’ forms.

Two-way random ICCs (with 95 % confidence interval, CI) in SPSS 20 were calculated to determine interrater reliability for all case vignettes and all possible pairs of raters for both the DD-CGAS and the OSU Autism CGI-S, as well as separately for experienced and less experienced clinicians. The ICC classification by Landis and Koch (1977) was used to interpret the findings, with ICC  $<.20$  indicating slight,  $.21-.40$  fair,  $.41-.60$  moderate,  $.61-.80$  substantial, and  $.81-1.00$  (almost) perfect agreement. In addition to ICCs, Pearson correlation between DD-CGAS and the OSU Autism CGI ratings was computed. In terms of construct validity, we expected a high negative

**Table 1** Case vignettes

No.	Age (years)	Gender	ASD diagnosis	IQ	Comorbidity	Treatment
1	15	Male	PDD-NOS	108	ADHD, anxiety, depression	CBT
2	11	Female	Asperger syndrome	99	ADHD, anxiety, sleep disorder, conduct disorders	CBT, pharmacological, family therapy
3	16	Male	Autism	75	DCD	Psychoeducation, CBT
4	8	Male	Asperger syndrome	106	ADHD, OCD	Social skills training, CBT
5	15	Female	Asperger syndrome	91	ADD, depression	CBT, social skills training
6	8	Male	Autism	76	ADHD, depression	CBT, psychoeducation
7	15	Male	PDD-NOS	95	Anxiety	CBT, psychoeducation
8	14	Female	Asperger syndrome	120	ADHD	CBT, psychoeducation, social skills training

*PDD NOS* pervasive developmental disorder-not otherwise specified, *CBT* cognitive behavior therapy, *OCD* obsessive compulsive disorder, *ADHD* attention-deficit hyperactivity disorder, *DCD* developmental coordination disorder

correlation between the DD-CGAS (with increasing values indicating higher adaptive functional skills), on one hand, and the OSU Autism CGI-S (with increasing values indicating higher symptom severity), on the other.

## Results

Average scores on the DD-CGAS made by experienced raters were  $M = 56.1$  at referral and  $M = 59.7$  at discharge. For inexperienced raters the figures were  $M = 51.1$  at referral and  $M = 58.6$  at discharge. On the OSU Autism CGI, experienced raters scored on average  $M = 3.4$  at referral and  $M = 3.0$  at discharge. In experienced raters scores on the OSU Autism CGI were  $M = 4.2$  at referral and  $M = 3.7$  at discharge. *ICCs* for all raters (experienced and inexperienced) and points in time were .63 (95 %  $CI = .34-.94$ ) on the DD-CGAS and .60 (95 %  $CI = .31-.93$ ) on the OSU Autism CGI-S. On the DD-CGAS, *ICCs* were 0.75 [95 %  $CI = .45-.96$ , (.78 referral/.79 discharge)] for experienced, and .58 (95 %  $CI = .39-.96$ , [ICC.54 referral/.53 discharge]) for inexperienced clinicians. On the OSU Autism CGI-S, *ICCs* were .72 [95 %  $CI = .39-.96$ , [ICC.73 referral/.72 discharge]] for experienced and .59 (95 %  $CI = .40-.79$ , [ICC.48 referral/.46 discharge]) for inexperienced clinicians. The correlations between the DD-CGAS and the OSU Autism CGI-S were  $r = -.86$  (referral), and  $r = -.82$  (discharge), respectively.

## Discussion

Economic, clinician-based, reliable measures of symptomatic and functional change are scarce in ASD. This this

study examined the interrater reliability of the of the DD-CGAS and OSU Autism CGI-S, two brief and clinician friendly tools to assess psychosocial functioning and clinical severity that only require a few minutes to complete, in clinicians with varying professional background and clinical experience, being rather naïve to the instruments. Our results endorse the usage of both instruments even in untrained clinicians. Interpersonal agreement was substantial for experienced and moderate even for inexperienced raters in ASD cases of varying clinical complexity. Prior inter-rater training and the scoring true cases rather than vignettes might further increase interpersonal agreement (Dyrborg et al. 2000; Lundh et al. 2010). Therefore, the reliability values reported here might rather be pessimistic than optimistic estimates of the instruments' fidelity. In addition, as hypothesized, DD-CGAS and OSU Autism CGI-S values correlated highly negative, indicating divergent validity. Findings are in line with earlier studies on the DD-GAS in pervasive developmental disorders, and the classical GAS version in child mental health samples. However, the current study markedly enlarges previous evidence for the DD-GAS in ASD (Wagner et al. 2007) to ASD cases with various comorbidity, European culture, and untrained clinicians with differing professional background and clinical experience. To the author's best knowledge, no psychometric studies on the OSU Autism CGI have been published yet. Thus, the current article is the first to provide particularly evidence for usage of this tool in ASD.

This study has several limitations, among them its rather narrow focus on interrater reliability, with no other aspect of reliability or validity being included. Moreover, as the present study is embedded into an ongoing large randomized controlled trial on group-based social skills training (NCT01854346), the vignettes did in the majority describe

intermediate to high functioning children and adolescents with ASD, not profoundly impaired and intellectually disabled cases, which hampers the finding's generalizability to the severe autism spectrum. Finally, observable change in the vignettes was moderate, and despite this may be characteristic of ASD over a shorter period of time, the sensitivity of the DD-CGAS and OSU Autism CGI-I to change could therefore only be demonstrated to a certain degree.

In conclusion, both the DD-CGAS and OSU Autism CGI-S are feasible tools for quick and general assessment of symptom severity and psychosocial functioning in ASD. They might hold promise for a range of purposes in outcome evaluation and observation of change, for instance in the scope of large scale quality and research registries, where quick and intuitive scales are essential. Such registries are widely used in Scandinavia and contribute important results to the study of ASD (e.g. Abel et al. 2013; Magnusson et al. 2013). A new medical records based nationwide research and quality registry for neurodevelopment disorders (“NEUROPSYK”; including ASD) is under construction in Sweden, where different variants of the CGAS (e.g. DD-CGAS) and CGI-S (e.g. OSU Autism CGI-S) are planned as mandatory outcome tools. Establishing an interrater reliability protocol for the DD-CGAS and OSU Autism CGI-S and other instruments for clinicians and researchers contributing to the registry is an essential part of this effort. For the latter purpose, the current findings offer valuable information.

## References

- Abel, K. M., Dalman, C., Svensson, A. C., Susser, E., Dal, H., Idring, S., et al. (2013). Deviance in fetal growth and risk of autism spectrum disorder. *American Journal of Psychiatry*, *170*, 391–398.
- APA. (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5™)*. Washington, DC: American Psychiatric Publishing.
- Arnold, L. E., Aman, M. G., Martin, A., Collier-Crespin, A., Vitiello, B., Tierney, E., et al. (2000). Assessment in multisite randomized clinical trials of patients with autistic disorder: the Autism RUPP Network. *Research Units on Pediatric Psychopharmacology. Journal of Autism and Developmental Disorders*, *30*, 99–111.
- Bölte, S. (2012). Brief report: The Social Responsiveness Scale for Adults (SRS-A): initial results in a German cohort. *Journal of Autism and Developmental Disorders*, *42*, 1998–1999.
- Bölte, S., Poustka, F., & Constantino, J. N. (2008). Assessing autistic traits: Cross-cultural validation of the social responsiveness scale (SRS). *Autism Research*, *1*, 354–363.
- Charman, T., & Gotham, K. (2013). Measurement issues: Screening and diagnostic instruments for autism spectrum disorders—lessons from research and practice. *Child Adolesc Ment Health*, *18*, 52–63.
- Constantino, J. N., & Frazier, T. W. (2013). Commentary: The observed association between autistic severity measured by the Social Responsiveness Scale (SRS) and general psychopathology: A response to Hus et al. *Journal of Child Psychology and Psychiatry*, *54*, 695–697.
- Constantino, J. N., & Gruber, C. P. (2012). *Social Responsiveness Scale* (2nd ed.). Los Angeles, CA: Western Psychological Services.
- de Bildt, A., Oosterling, I. J., van Lang, N. D., Sytema, S., Minderaa, R. B., van Engeland, H., et al. (2011). Standardized ADOS scores: measuring severity of autism spectrum disorders in a Dutch sample. *Journal of Autism and Developmental Disorders*, *41*, 311–319.
- Dyrborg, J., Larsen, F. W., Nielsen, S., Byman, J., Nielsen, B. B., & Gautre-Delay, F. (2000). The Children's Global Assessment Scale (CGAS) and Global Assessment of Psychosocial Disability (GAPD) in clinical practice—substance and reliability as judged by intraclass correlations. *European Child and Adolescent Psychiatry*, *9*, 195–201.
- Gotham, K., Pickles, A., & Lord, C. (2012). Trajectories of autism severity in children using standardized ADOS scores. *Pediatrics*, *130*, e1278–e1284.
- Guy, W. (Ed.). (1976). *ECDEU assessment manual for psychopharmacology* (pp. 218–222). Rockville, MD: US Department of Health, Education, and Welfare Public Health Service Alcohol, Drug Abuse, and Mental Health Administration.
- Howlin, P., Magiati, I., & Charman, T. (2009). Systematic review of early intensive behavioral interventions for children with autism. *American Journal on Intellectual and Developmental Disabilities*, *114*, 23–41.
- Howlin, P., & Moss, P. (2012). Adults with autism spectrum disorders. *Canadian Journal of Psychiatry*, *57*, 275–283.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.
- Lord, C., Carr, T., & Grzadzinski, R. (2013). Measuring emerging changes in social communication. *Educational symposia at the International Meeting for Autism Research*, San Sebastian.
- Lord, C., Luyster, R., Guthrie, W., & Pickles, A. (2012). Patterns of developmental trajectories in toddlers with autism spectrum disorder. *Journal of Consulting and Clinical Psychology*, *80*, 477–489.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*, 659–685.
- Lundh, A., Kowalski, J., Sundberg, C. J., Gumpert, C., & Landen, M. (2010). Children's Global Assessment Scale (CGAS) in a naturalistic clinical setting: Inter-rater reliability and comparison with expert ratings. *Psychiatry Research*, *177*, 206–210.
- Lundh, A., Kowalski, J., Sundberg, C. J., & Landén, M. (2012). A comparison of seminar and computer based training on the accuracy and reliability of raters using the Children's Global Assessment Scale (CGAS). *Administration and Policy In Mental Health*, *39*, 458–465.
- Magnusson, P. K., Almqvist, C., Rahman, I., Ganna, A., Viktorin, A., Walum, H., et al. (2013). The Swedish Twin Registry: establishment of a biobank and other recent developments. *Twin Research and Human Genetics*, *16*, 317–329.
- Oono, I. P., Honey, E. J., & McConachie, H. (2013). Parent-mediated early intervention for young children with autism spectrum disorders (ASD). *Cochrane Database of Systematic Reviews*, *4*, CD009774.
- Reichow, B., Steiner, A.M., Volkmar, F. (2012). Social skills groups for people aged 6 to 21 with autism. *Cochrane Database of Systematic Reviews 2012*, Issue 7. Art. No.: CD008511.

- Rutter, M., LeCouteur, A., & Lord, C. (2003). *Autism Diagnostic Interview-Revised (ADI-R)*. Los Angeles: Western Psychological Services.
- Shaffer, D., Gould, M. S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., et al. (1983). Children's global assessment scale (CGAS). *Archives of General Psychiatry*, *40*, 1228–1231.
- The OSU Research Unit on Pediatric Psychopharmacology. (2005). OSU Autism Rating Scale. Columbus, OH: Author. © by OSU Research Unit on Pediatric Psychopharmacology.
- Tse, J., Strulovitch, J., Tagalakis, V., Meng, L., & Fombonne, E. (2007). Social skills training for adolescents with Asperger syndrome and high-functioning autism. *Journal of Autism and Developmental Disorders*, *37*, 1960–1968.
- Wagner, A., Lecavalier, L., Arnold, L. E., Aman, M. G., Scahill, L., Stigler, K. A., et al. (2007). Developmental disabilities modification of the Children's Global Assessment Scale. *Biological Psychiatry*, *61*, 504–511.
- Wood, J. J., Drahota, A., Sze, K., Van Dyke, M., Decker, K., Fujii, C., et al. (2009). Brief report: effects of cognitive behavioral therapy on parent-reported autism symptoms in school-age children with high-functioning autism. *Journal of Autism and Developmental Disorders*, *39*, 1608–1612.
- World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and guidelines*. Geneva: WHO.