Applied Behavior Analysis and the Autism Diagnostic Observation Schedule (ADOS): a Symbiotic Relationship for Advancements in Services for Individuals with Autism Spectrum Disorders (ASDs)

John M. Guercio · Adam D. Hahs

Published online: 6 January 2015
© Association for Behavior Analysis International 2014

Abstract The Autism Diagnostic Observation Schedule (ADOS) is a widely used diagnostic tool for the assessment of autism spectrum disorders. We compared interobserver reliability scores for observers using the existing guidelines for administration of the ADOS with those of the same observers using modified guidelines that included operational definitions for each area of observation. The results indicated increased mean interobserver reliability scores from 55.83 % using the existing guidelines to 84 % when using the modified guidelines and enhanced reliability scores for 3 of 3 clients with whom the ADOS was used. The guidelines are the following: integration of a highly effective diagnostic tool with ABA, the inclusion of a diagnostic tool may promote more comprehensive ABA services, more clearly define and identify areas of concern in autism, and impact implementation efficacy and reliability of assessments for individuals with autism.

Keywords Autism spectrum disorders · The autism diagnostic observation schedule (ADOS) · Applied behavior analysis

The diagnosis of autism spectrum disorders (ASD) is an area that has posed problems for clinicians (Tomanik et al. 2007). The primary issue was that ASDs share some common behavioral characteristics with other disorders such as attention deficit disorder (ADD) or intellectual disability (ID). The expression of the symptoms inherent in autism may manifest differently across children as they differ with

respect to where they fall on the spectrum. The qualitative impairments in communication, social interaction, and the display of repetitive behaviors are present in the autism population, but may be observed in differing frequencies or durations (Gillham, Carter, Volkmar, & Sparrow, 2000). The initial screening of individuals for ASDs is typically done using the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 1989).

The ADOS is a semi-structured assessment used to evaluate the social, play, and imaginative repertoires (Lord et al., 1989) of children with autism. It was first introduced in the late 1980s and was targeted for children with language abilities around 36 months (Lord and Risi 1998). It is comprised of 4 assessment modules, with the intent that they be completed via direct observation of participants. The assessment tool can be utilized across a variety of age groups and developmental and language abilities (Lord and Risi 1998). Individual items are scored using a 4-point scale. In addition to other clinical assessment tools, the ADOS is used both as a diagnostic tool and a screening assessment to determine which services and programs may be most beneficial to the client. If the individual's behavior for a specific skill was atypical and created a barrier for their functioning, a score of 3 would be given (Akshoomoff et al. 2006).

When completing the ADOS, each item is scored based on the administrator's direct observations of client's performance during the assessment. Given the human element involved in scoring, examiners may disagree on the rating to assign a particular client on that section. A diagnostic impression is very important when it comes to a determination of whether or not a child has signs or symptoms indicative of an autism spectrum disorder. Therefore, the success of any rater using the ADOS is dependent upon the skill of the examiner (Lord and Risi 1998). Most of the tests that have been conducted on

J. M. Guercio (✉) · A. D. Hahs
Benchmark Human Services & Arizona State University, St. Louis & Tempe, MO & AZ, USA
e-mail: jguercio@awsusa.com

the interobserver reliability of the ADOS have been conducted with examiners that have been trained to administer the instrument (Lord and Risi 1998). While training in the ADOS would help reliability, providing operational definitions of each of the subsections may allow the assessments to be scored with even greater accuracy. The inclusion of operational definitions might also have implications for ADOS training, as it might be possible to reduce the amount of training required if clinicians were more accurate while scoring the assessment. Thus, the purpose of the current investigation was to examine the effect of modifying the scoring criteria to include operational definitions on interobserver reliability.

Using the conventional ADOS scoring guidelines, the mean IOR percentage was 55.83 %. With the inclusion of the operational definitions to the ADOS scoring, the mean IOR percentage was 84 %. Therefore, there was a 28.17 % increase with the modified scoring system from the conventional guidelines. Figure 1 depicts the differences between the conventional scoring approach and that of the use of the operationally defined criteria. We conducted statistical analyses for these differences using SPSS, and a statistically significant difference was found between the mean of the IOR scorings using the original ADOS ($M=55.83$, $SD=11.16$) and those of observers using the operationally defined ADOS ($M=84.00$, $SD=5.76$); $t(5)=-8.94$, $p\leq.05$. Thus, the inclusion of the operational definitions increased reliability between observers and increased that reliability across all ADOS modules.

Figure 2 represents the mean IOR differences between scorings of individual clients when using the conventional scoring criteria. Percentage differences in IOR means were 36, 18, and 30 % for Hayden, Ryan, and Hunter, respectively, in favor of the operational definition criteria. These differences also yielded statistically significant differences for each of the three individual's IOR means at $p\leq.05$. An interesting finding to note from Fig. 2 is that the current results indicate not only higher levels of IOR, but higher levels of IOR regardless of the client's chronological or developmental age.

Fig. 1 ADOS sections from modules 1, 2, 3, and 4 along with their corresponding interobserver reliability scores. The *black bars* represent scoring with the standard definitions, and the *white bars* represent scoring using operational definitions

Fig. 2 Mean interobserver reliability scores for the clients. The *black bars* represent scoring with the standard definitions and the *white bars* represent scoring using operational definitions

Each pair of observers viewed tapes from each of the modules that the ADOS offers. After each of these viewings using the standardized ADOS definitions, the pair then re-viewed the tape and scored it using the operational definitions. One possible limitation of the study was the order in which observers scored the modules (i.e., without operational definitions then with operational definitions), suggesting possible practice effects. However, the fact that successive modules of the ADOS were completed with the same results being shown every time seems to diminish the role that practice effects could have had on the outcomes observed in the study. Increases in interobserver reliability were always seen when the operational definitions were being employed as opposed to the existing or conventional definitions. This pattern repeated several times through the course of the study, indicating that the observed results were most likely due to increases in specificity given the use of the operational definitions as opposed to practice effects. The fact that these definitions were provided to the observers after they had been given the opportunity to observe and score the ADOS using the standard definitions shows the utility of the inclusion of the definitions. Future studies could examine the role of the order of observations. The existing ADOS definitions and the operational definitions could be rotated in terms of which one the staff used first in their scoring of the videos.

Another fact to be taken into account is the increased ease of obtaining higher interobserver reliability scores with behaviors that can be easily measured. The stereotyped and repetitive behavior section of the ADOS seems to be the most amenable to the addition of operational definitions to enhance the accuracy of interobserver reliability in this section. The ability to accurately screen for autism is extremely important for a number of reasons. Parents can gain better access to services once a diagnosis is made. The data in this study are quite demonstrative that we have the capacity to make these changes; the question remains as to how willing we are to implement it.

Methods

All sessions took place in the offices of a treatment center for individuals diagnosed with autism. Two clinical staff and two graduate interns served as observers who scored ADOS assessments. Observer 1 was a Board Certified Behavior Analyst (BCBA) and observer 2 was a certified occupational therapist assistant. Both observers 1 and 2 were employed as clinicians at the treatment center. Observers 3 and 4 were two behavior analysis and therapy interns completing assistantships at the center. None of the individuals observing and scoring the assessments had previous training with respect to the application and interpretation of the assessment. All videotaped assessments were conducted in a room equipped with a video camera and a two-way observation mirror. Each room contained a table and three chairs. The ADOS was administered by two clinicians, Liz and Amy, both of whom had previously completed the UMACC ADOS training workshops. Six male clients participated in the assessments. Client ages ranged from 3 years to early twenties. The ADOS was administered to each client primarily for diagnostic purposes and treatment recommendations.

A total of six ADOS assessments were scored during baseline. Observers 1 and 2 independently observed and scored two, module 1 and one, module 3 assessments. Observers 3 and 4 independently observed and scored one module 2, one module 3, and one module 4 assessments. Each of the observers scored three ADOS assessments from client videotape using the coding definitions in the conventional

ADOS test booklets. Of the ADOS assessments selected for scoring, three were completed by Liz, and three were completed by Amy. In order to account for any possible biases, the assessments for each module were selected randomly from a collection of tapes that each clinician kept of their client sessions (Table 1).

Prior to viewing any tapes, the BCBA and behavior analysis and therapy interns developed operational definitions for each coding item for modules 1–4 of the ADOS. Operational definitions provide individuals who may be recording data or observing some behavior, specific guidelines as to what denotes an occurrence of the targeted behavior by specifying the dimensions of the behavior (i.e., frequency, duration, intensity). For purposes of the current experiment, operational definitions were developed by adding frequency counts, percentage measures, and observable descriptions of target behaviors to the conventional ADOS coding system descriptions used with the assessment. For example, the existing ADOS scoring definitions for “Immediate Echolalia” for module 1 read: 0=rarely or never repeat others’ speech, 1=occasional echoing, 2=echoing words and phrases regularly, but some spontaneous language (it can be stereotyped), 3=speech largely consists of immediate echolalia, 8=language too limited to judge. The same subset with the operational definitions included reads: 0=zero instances of repeating others’ speech, 1=25–50 % of total speech consists of repeating others’ speech, 2=50–75 % of total speech consists of repeating others’ speech, 3=75 % or more of total speech consists of repeating others’ speech,

Table 1 ADOS sections from modules 3 and 4 which are used to assess higher functioning individuals are presented in the first column on both the upper and lower panels. The second panel contains sample definitions

as they appear in the test booklet published by Western Psychological Associates. The third column in each panel includes the operational definitions for each sample definition presented in the second column

Module 3		
ADOS module 3 section	ADOS module 3 definition	Operational definition
Stereotyped/Idiosyncratic use of words or phrases	<i>Often</i> uses stereotyped utterances or odd words or phrases, either with or without language.	Uses stereotyped utterances or odd words or phrases at <i>least five times</i>
Reciprocal social interaction	Uses <i>poorly modulated</i> eye contact to initiate, terminate, or regulate social interaction.	Uses eye contact in durations of <i>less than 2 s each</i> to initiate, terminate, or regulate social interaction during at least 70 % of the session.
Quality of social response	<i>Odd, stereotyped responses</i> , or responses that are restricted in range or inappropriate to the context.	<i>More than 50 %</i> of the social responses are stereotyped or are inappropriate in context
Hand and finger and other complex mannerisms	<i>Very brief or rare</i> hand and finger mannerisms or complex mannerisms, OR mannerisms not as clear as specified below for a rating of 2.	Very brief (<i>less than 3 s</i>) or rare (<i>fewer than 3 instances</i>) hand and finger mannerisms or complex mannerisms
Module 4		
ADOS module 4 section	ADOS module 4 definition	Operational definition
Descriptive, conventional, instrumental, or informational gestures.	Some spontaneous use of informational, conventional, or instrumental gestures, but not descriptive.	Spontaneous use of AT LEAST 2 gestures that are used in communicating with the examiner.
Emphatic or emotional gestures	Odd, excessive, or definitely awkwardly integrated emphatic or emotional gestures.	At least 3 odd or awkwardly integrated emphatic or emotional gestures

Words highlighted in italics were identified as the key properties of each definition. The italicized words in the second column were deemed to be somewhat ambiguous and were replaced with the descriptors in bold in column

8=language too limited to judge. Similar modifications were made to each of the coding descriptions within each module. Observers independently re-viewed the same videotapes that they had in baseline and rescored the same assessment modules using the included operational definitions for the ADOS.

Once the four observers had completed scoring the assessments from videotape in both conditions, interobserver reliability (IOR) was calculated between the observers for each of the individuals in each module. A mean IOR percentage for the module was then calculated via averaging the two individual IOR percentages in each module. The aforementioned steps for IOR calculation were then repeated for condition 2. Thus, there were 8 resulting interobserver reliability percentages (i.e., condition 1 and condition 2 IOR percentages for each of the four modules). IOR was calculated by dividing the number of agreements by the sum of agreements plus disagreements and multiplying the resulting quotient by 100 %.

References

- Akshoomoff, N., Corsello, C., & Schmidt, H. (2006). The role of the autism diagnostic observation schedule in the assessment of autism spectrum disorders in school and community settings. *California School Psychologist, 11*, 7–19.
- Gillham, J., Carter, A., Volkmar, F., & Sparrow, S. (2000). Toward a developmental operational definition of autism. *Journal of Autism and Developmental Disabilities, 30*, 269–288.
- Lord, C., & Risi, S. (1998). Frameworks and methods in diagnosing autism spectrum disorders. *Mental Retardation and Developmental Disabilities Research Reviews, 4*, 90–96.
- Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., et al. (1989). Autism diagnostic observation schedule: a standardized observation of communicative social behavior. *Journal of Autism and Developmental Disorders, 19*, 185–212.
- Tomanik, S. T., Pearson, D. A., Loveland, K. A., Lane, D. M., & Shaw, J. B. (2007). Improving the reliability of autism diagnosis: examining the utility of adaptive behavior. *Journal of Autism and Developmental Disabilities, 37*, 921–928.