

An Evaluation and Critique of the Childhood Autism Rating Scale: Second Edition (CARS2)

Chapman University

An Evaluation and Critique of the Childhood Autism Rating Scale 2 (CARS2)

Description

The *Childhood Autism Rating Scale-Second Edition (CARS2)* builds on the *Childhood Autism Rating Scale* published in 1988. CARS2 is authored by Eric Schopler, Ph.D., Mary E. Van Bourgondien, Ph.D., G. Janette Wellman, Ph.D., and Steven R. Love, Ph.D. and was published in 2010 by the Western Psychological Services. The CARS2 consists of three forms to identify the presence of autism spectrum disorder (ASD) symptoms as well as to discriminate between a child with autism and a child with other cognitive deficits. It also distinguishes children with ASD in the moderate-to-moderate range from children in the moderate-to-severe range. The three forms are the standard version (CARS2-ST), the Questionnaire for Parents or Caregivers (CARS2-QPC), and the high functioning version (CARS2-HF) which is intended for children with fluent communication skills who are six or older with an IQ above 80. The CARS2-ST is the same as the original CARS. The CARS2-ST should be used for children with an IQ of 79 or lower, who are younger than six years old and have impaired communication. The addition of CARS2-HF allows for a wider use of the CARS and is a strength of the updated version.

Content and Use

Users and Uses of the CARS2

The information collected from the CARS2 provides a summary of information that can be used to develop a diagnostic hypothesis among individuals who are suspected of falling within the ASD category. Ratings on the CARS2-ST and CARS2-HF are based on frequency, intensity, peculiarity, and duration of behaviors. Professionals can use data collected from the CARS2 to give diagnostic feedback to parents and guide interventions. The CARS2 is an

untimed observational assessment and was designed to be used in clinics, schools, and intervention programs provided the test administrator has training with individuals with autism. Level C qualification (master's or bachelor's degree in psychology, school counseling, occupational therapy, speech-language pathology, social work, education, special education or a related field) is required to administer the CARS2-ST and CARS2-HF tests. Since a variety of professionals who interact with a child can administer the test, the responses are subjective. Administrators of the test may have different perceptions of normal behaviors leading to rater error. The use of the CARS2 must remain within the examiner's scope of practice; a school psychologist may use results to determine placement but cannot give a medical diagnosis. It is important to understand the CARS2 as a piece of an assessment when diagnosing ASD and should not be used alone to produce a diagnosis.

A key difference between the administering of CARS2-ST and the CARS2-HF, is the CARS2-HF requires the rater to receive information from multiple sources. It is necessary to interview an individual who is familiar with the child's behavior in different settings along with the examiner's direct observation of the child's behaviors. The CARS2 –ST ratings may come from a single source such as observation, psychological testing, classroom participation, parent report, or comprehensive clinical records. Time, setting, source, and method error should be considered due to the variety of conditions and settings in which observations for the CARS2 can take place. The CARS2 manual costs \$88.00. The CARS2-ST and CARS2-HF rating booklets come in packs of 25 and cost \$47.00, and the CARS2-QPC costs \$33.00 for a pack of 25.

Characteristics of the Test

The CARS2-QPC is a data collection tool used in conjunction with the CARS2-ST or CARS2-HF. The CARS2-QPC asks parents or caregivers to rate the child on the same 15

dimensions of the CARS2. There is a space on each page for parents to write out specific examples. The CARS2-QPR was made to facilitate the clinical process and assist in distinguishing between mid-to-moderate and severe ASD.

Both the CARS2-ST and CARS2-HF are made up of 15 items that corresponds to a specific domain of functioning related to ASD (i.e. relating to people, imitation, emotional response, body use, object use, listening response, taste/smell/touch response and use, fear/nervousness, verbal communication, nonverbal communication, activity level, level/consistency of intellectual response, and general impressions). Examiners rate examinees on a rating scale of 1-4 as well as write in examples of each behavior. While rating the examiner must also take peculiarity, frequency, intensity, and duration of the child's behaviors into consideration. The rating values for each item are as follows:

- 1 Within normal limits for that age
- 1.5 Very mildly abnormal for that age
- 2 Mildly abnormal for that age
- 2.5 Mildly-to-moderately abnormal for that age
- 3 Moderately abnormal for that age
- 3.5 Moderately-to-severely abnormal for that age
- 4 Severely abnormal for that age

These items are subjective ratings and can be inconsistent when considering the wide range of professionals who may be administering the test. Individuals may have different opinions of what "level and consistency of intellectual response" means as well as what constitutes as abnormal for the child age. Ratings based on parent reports are more likely to be skewed, as they want their child to look better. Parents may also lack experience with other

children which may cause them to view their child's abnormal behavior as more typical or typical behavior as abnormal.

To combat rater's differences in perceptions, surround a child's behavior the rating booklet provides a comprehensive, detailed item-specific rating instructions. These descriptions include example behaviors represented by each rating level. For example, the explanation of a rating of "1" under "Body Use" on the CARS2-HF states the child "moves with the same ease, agility, and coordination as a typical person of the same age"; while a rating of "3" states the child "displays any body posture or stance, hand or fingers mannerism, flapping, self-directed aggression, picking at body rocking, spinning, or toe-walking. Fine motor difficulties or obvious handwriting difficulties are present, which may result in resistance to writing tasks". Rating values are summed to produce a total raw score which is converted to a *T* score and corresponding percentile rank. The CARS2-ST and CARS2-HF provide a cutoff score to determine whether examiners need to evaluate the presence of autism further.

Each section of the CARS2-ST and CARS2-HF provides a space for note-taking and written examples of behaviors observed for each category. The authors decided to add in a note-taking section in the updated version to make it more user-friendly and comprehensive. I believe the CARS2-HF booklet provides enough space for observations and notes in regards to the child's behavior. The detailed rating descriptions are helpful for raters; yet, the descriptions are in a small font and can be difficult to read. The scoring section of the rating booklet is simple and easy to follow. The front of the book contains a raw score table that converts scores to a *T*-score and percentile rank. The standard scores reflect a continuum of behaviors related to ASD. A box is provided showing where scores fall on a range of behaviors from minimal-to-no

symptoms of ASD to severe symptoms of ASD. It is important to note the CARS2-ST items and cut off standard scores have not been changed from the original version.

The CARS2-QPC contains seven sections. Sections one through six include statements for parents/caregivers to rate the child's functioning on a 5 point Likert scale. The seventh section consists of open-ended questions about behaviors not covered in the previous sections, and the parent/caregiver is asked to write out their responses.

Characteristics of the Test Manual

The CARS2 rater manual is well organized, clean, and comprehensive. The authors provided a clear introduction to the test, description, and explanations regarding the intended uses and users of the test. The manual is divided into two parts. The first part consists of five chapters which address administering, scoring, how to make observations and ratings, case examples, interpretations of scores, and intervention and planning resources. Part two of the manual consists of two chapters addressing psychometric properties of the test, the development of the CARS method, rationale for rating items, and the relationship of CARS to CARS2-ST, CARS2-HF, and other diagnostic criteria.

The chapter regarding scoring goes into depth of what each rating value looks like in child's behaviors. Descriptions of behaviors provided in the rating booklet are to be used as a guide for rating behaviors; therefore, raters should closely study the item descriptions and rating criteria in the test manual found under the section *How to Make CARS2-ST Observations and Ratings* in chapters two and three.

I found the case studies provided in the manual to be exceptionally helpful. A case example was given for the CARS2-ST and CARS2-HF. The authors provided an example of a scored rating booklet for each case and applied the scores for examples of diagnosis and

intervention planning. The case studies are helpful to understand better how a child's observed behaviors relate to the scales provided in the rating booklet. The fifth chapter focuses on intervention planning. The authors provide specific intervention strategies, planning, and resources. The information provided does not cover all possible strategies or available resources but serves as a starting point for professionals who may be new to understanding the many features of ASD. The authors provide relevant resources that address the wide variety of domains of behaviors presented in ASD as well as diving resources and interventions by the CARS2-ST and the CARS2-HF.

The manual also contains defined sections explaining reliability and validity. The explanations provided are comprehensive and understandable for individuals who are not familiar with the terms reliability and validity. These sections also include organized tables and figures which are helpful visuals. In sum, I found the test manual to be user-friendly and accessible to professionals with various levels of training.

Standardization and Sample Norms

The norming sample for both the CARS2-ST and CARS2-HF are composed only of individuals with diagnosed ASD. Scores on the CARS2 reflect the functioning of the ASD population rather than typically developing children. The CARS consist of a large total sample of over 3,500 children and adults who participated in the standardization of the test; the original CARS2 development sample (N=1,606), the current CARS2-ST verification sample (N=1,034), and the CARS-HF developmental sample (N=994). The test manual describes the original CARS sample in detail. Those in the verification CARS2-ST sample were individuals referred to one of five Division TEACHH centers over the span of 15 years; but, the states of the

TEACHH centers was not provided. The CARS2 was normed based on data from the U.S. Census of 2000.

CARS2-ST Verification Sample

The CARS2-ST verification sample reflected the demographic characteristics of the 2000 Census and differed from the sample norm of the original CARS. More than half of the verification sample consisted of males (N=804; 78%) which reflects data that more male individuals have ASD. The ages of individuals ranged from two to 36 years, with 73% of individuals in the verification sample being under 11 years old. The manual's language refers to children rather than adolescents or adults. The CARS2 is most frequently used with young children; however, it does not discourage its use with adolescents and adults. The language of the rating booklet is not user-friendly when observing adolescents and adults, and the examiner should take this as well as the verification sample characteristics into consideration. The ethnic background of the normative sample closely reflected the 2000 census with 60% of the sample described as White, 16% described as Black/African American, 13% Hispanic/Latino, 7% Asian/Pacific Islander, and 4% Other. The IQ of the verification sample group is under 85, with 81% of the group equal to or under an IQ of 79. Full-scale IQ estimates were based on standard measures such as the WISC-IV, the *Stanford-Binet* (Roid, 2003), and the *Test of Nonverbal Intelligence* (Brown, Sherbenou, & Johnsen, 1997). This skewed sample is appropriate since the CARS2-ST is designed for those with low cognitive functioning and is representative of the children who will be administered the test. The manual provides information regarding years of education completed by the head of household, as it is a strong indicator of socio-economic status. However, the gender of the head of household was not provided. Geographic region of the sample was broken up into the Northwest, Midwest, South, and West. The verification

sample reflected the 2000 Census populations, though the participant percentage from the Midwest was 12% lower than the US population in the 2000 census. Information regarding the number of participants from individual states was not provided.

CARS2-HF Development Sample

The individuals rated in this sample presented a variety of clinical diagnoses including, high functioning autism ($n = 248$), Asperger's Disorder ($n = 231$), ADHD ($n = 179$), Learning Disorder ($n = 111$), and other clinical disorders ($n = 69$). A small group of students in general education were included ($n = 21$) as well as non-autistic students in special education ($n = 40$) to verify an absence of symptoms rated on the CARS2-HF. Demographic characteristics of this sample were similar to the CARS2-ST sample. Seventy-six percent of individuals were under age fifteen, and 78% were male. Ethnicity and head of household-years of education completed of the development sample are reflective of the 2000 census report. The development sample overrepresented the Southern region by 17% and lacked participants from the Western region; the West made up 23% of the US population in the 2000 census and only 10% of the development sample. All participants had an IQ estimate of 80 or higher since the CARS2-HF is designed for individuals with average or higher cognitive abilities. The DSM-5 removed Asperger's syndrome as a diagnostic category and should be reflected in the next revision of the CARS manual.

Scores and Interpretation

When observing and rating an individual the manual addresses the individual's chronological age, peculiarity, frequency, intensity, and duration of behaviors. The peculiarity is defined as the extent to which behavior is unusual or odd as opposed to whether it is delayed or similar to the behavior of a younger child. Although this definition is helpful, the measure is

subjective and could be influenced by the rater's perception of "odd behaviors." Each item must be rated to obtain a measurable score for interpretation. Any single source of information, such as psychological testing, classroom participation, parent reports, and comprehensive medical records can be used to provide information regarding the 15 items. Observations of a child can be narrative or structured. Structured observations, typically in clinical settings, may be problematic when observation procedures are not followed across observations/ observers. There is room for possible interpretation error of the CARS2 due to its ability to be used in a wide range of settings and by a wide variety of professionals.

The CARS2 rating booklet is lightweight and well formatted with criteria and definitions extensively laid out, though, the definitions of criteria are written in a small font, and the rater may be distracted reading the booklet rather than observing the child. Scoring the rating booklet is simple. The ratings of all 15 categories are transferred to one page. The manual describes the process of scoring as well as provides an example of a completed CARS2 rating booklet.

There are three types of score results, total raw score, *T*-score, and rating value for each item. The CARS2-ST cut off value was based on the comparison scores of corresponding clinical assessments of over 1,500 children. Children under 13 years with scores lower than 29.5 and adolescents/adults 13 or older with scores of 27.5 or lower are categorized as likely not to have ASD. Children with a score of 30 or above and adolescents/adults with a score of 28 or higher are categorized as likely to have ASD. The cutoff points given are meant for post-referral screenings only, and should not be applied to every individual. *T*-scores are useful to make comparative judgments regarding the severity level of ASD related behaviors present in an individual. The *T*-score values that correspond to the CARS-2 total raw scores have a mean of

50*T* and a standard deviation of 10*T*. These scores are based on a sample of over 2,000 individuals with autism, ages 2 to 57.

Psychometric Properties

Reliability

The authors used Cronbach's alpha coefficient to examine the internal consistency reliability of the CARS. The original CARS obtained an internal consistency estimate of .94 and the manual lists other authors who reported Alpha coefficient scores ranging from .73 to .94. The CARS2-ST and CARS2-HF both obtained high internal consistency estimates, .93 and .96, respectively. These Alpha scores indicate the items represent the domain of ASD. An internal consistency of .90 is sufficient for educational and clinical applications.

Interrater reliability examined whether different professionals would rate referred individuals similarly. The median of interrater reliability on the CARS was .71, ranging from .55-.93. The CARS2-HF interrater reliability estimates range from .53-.93 with a median correlation of .73. These scores are considered adequate; however, should be further examined. Raters were given clinical information regarding referred individuals, yet, in practice, the CARS2 is used as an observational measure. Furthermore, there is a lack of information regarding the trained raters. No information was provided regarding level and type of training, gender, setting the rating occurred, and the sources of information being utilized. A measure of interrater reliability was not provided for the CARS2-ST.

On the original CARS, ratings by professionals with limited experience as well as ratings by expert clinical directors were correlated, and scores ranged from .73 to .83. The authors state that correlations were obtained across "different types of raters and settings, using varying sources of information" (p.79). To examine interrater reliability of the CARS a sample of

professionals in a clinical setting, who were not experts in autism, were introduced the instrument and then asked to make ratings during their observations during diagnostic sessions. The professionals were given an hour to read over the CARS manual and viewed a 30-minute training tape. These scores were compared to clinical directors. The total scores obtained for visitors' ratings had a significant correlation with those obtained for ratings by clinical directors ($r = .83, p < .01$). No information was provided on what qualities constitutes someone as an "expert." The sample size was small and consisted of 18 individuals including medical students, pediatric residents, special educators, school psychologists, and speech pathologists. The authors imply the data from the CARS translates to the CARS2-ST and valid ratings on the CARS2-ST can be made by a variety of professionals who have minimal training or experiences with individuals with ASD.

Scores for two test occasions, approximately one year apart were compared for 91 cases in the CARS development sample to assess the stability of ratings over time. To avoid improvements frequently seen in behavior during the first a second assessment period following initial intensive interventions, total scores for the second and third yearly evaluations were chosen. The resulting correlation was $.88 (p < .01)$. This indicates the ratings on the CARS are stable over time, yet, it is important to note the small sample size of cases re-tested; 150 cases would be a better sample size for test-retest reliability. Literature supports the stability of scores. The manual cites Perry and Freeman (1996) to have found a retest correlation on $.90 (n = 11)$ one year apart and $.78 (n = 30)$ two years apart for children and adolescents receiving autism-related interventions.

To account for the possible effect of random variation due to measurement error, the *SEM* indicates the range in which a "true" score would likely fall (with a 68% probability), given

the obtained score. The *SEM* for the CARS2-ST total raw score is 0.68, and the *T*-score is 2.7*T*. This indicates that a “true” *T*-score, with a 68% probability, is within less than 3 *T*-score points above or below the obtained *T*-score. The *SEM* for the CARS2-HF total raw score is 0.73, and the *T*-score is 2.8 meaning the “true” *T*-score is probably within less than 2.8 *T*-score points above or below the obtained *T*-score. Although a 68% confidence level is frequently used for psychological tests, a higher confidence level would be ideal to interpret a score that considers measurement error more accurately. The manual provides steps for how to interpret scores with a higher confidence level.

Validity

Validity refers to the accuracy and appropriateness of the interpretation of a test score. Validity is an ongoing process and is evolved through the accumulation, evaluation, and synthesizing of research, theory, and concept development. The authors of the CARS2 list theory-based evidence, evidence based on the test content, information about the internal relationships of the item ratings and Total scores, and the relationship of item ratings and Total score to variables external to the instrument as common types of validity evidence important to instruments such as the CARS2. Internal validity was assessed for the CARS2-ST and the CARS2-HF. Correlations among item ratings and total raw scores were moderate ranging from .42 to .77 for the CARS2-ST and .40-.79 for the CARS2-HF.

Sensitivity refers to the test's ability to correctly identify individuals with a given disorder, whereas specificity refers to the test's ability to identify those without the disorder correctly. The sensitivity of the original CARS used a total raw score cutoff value of 30 which correctly identified 87% of individuals in the development sample as having or not having ASD. When using the total raw score, identification of individuals with or without ASD resulted in a

sensitivity of .88 and .87, respectively. Ratings for the CARS2-ST verification sample are consistent with the distribution of scores on the CARS development sample. Sensitivity and specificity values for the CARS2-HF were .81 and .87, respectively. These scores are sensitive for distinguishing high-functioning individuals with ASD from all other groups. Differences in rating patterns occurred between older and younger individuals with ASD as well as between individuals with Asperger's syndrome and individuals diagnosed with ASD. The average CARS2 total raw score obtained in the sample of 1,034 individuals with an autism diagnosis was 38.5 ($SD = 8.4$), which supports the established cutoff score of 30, or one SD below the mean for diagnosed ASD cases.

Criterion-related validity on the CARS2 was addressed by examining the relationship of total scores on the CARS2-ST and CARS2-HF to other tests that measure ASD. Correlations with the Autism Diagnostic Observation Schedule (ADOS; Lord & Corsello, 2005) were .79 for the CARS2-ST and .77 for the CARS2-HF. These scores indicate a strong relationship between clinician ratings. Independent ratings of autism symptoms obtained for the original CARS was .84, indicating a high validity when compared to clinical ratings. The correlation between the CARS2-ST and the Autism Behavior Checklist (ABC; Krug, Arick, & Almond, 1980) was .67. Eaves and Milner (1993) also examined the correlation between the CARS and the ABC and found a correlation of .72. The CARS correctly identified 98% of the autistic individuals; it identified 69% of the possibly autistic as autistic. The ABC correctly identified 88% of autistic individuals. Studies agree the CARS is an accurate assessment of ASD due to its relatively high concurrent validity. When compared to the DSM-IV researchers found the CARS to be highly correlated to clinical diagnosis as well as, having a sensitivity of .94 and a specificity of .85 (Perry, Condillac, Freeman, Dunn-Geier, & Belair, 2005). More recently, 97% of children who

were diagnosed as low-functioning were correctly identified by the CARS, while 75% of high functioning children were correctly identified (Dickerson, S. et al., 2009). The CARS2-HF responded to the need for a separate assessment for high functioning individuals with ASD. Researchers also found a 97% sensitivity rate and 100% specificity rate when children diagnosed with ADHD were assessed.

Conclusion

There are several advantages to the CARS2. First, the inclusion of the high-functioning version allows the CARS2 to be more inclusive and allows for higher sensitivity and specificity. The standardization samples were generally representative of the population it seeks to assess in accordance with the U.S. census of 2000. The CARS 2 total score is useful in understanding the continuous severities of ASD. The test demonstrates strong validity through its high correlations with other measures of ASD and high sensitivity and specificity values. A particularly useful aspect of the CARS2 is the inclusion of intervention planning and resources. The CARS2 provides tools professionals can use to inform their practice while working with individuals with varied levels of severity of ASD. The CARS2 has been translated into a variety of languages including Japanese, Swedish, Korean, Spanish, French, and Chinese which makes it more accessible to children and professionals from a variety of backgrounds. The CARS2 can be used in a variety of settings by a variety of professionals. Some limitations of the CARS2 include the use of the 68% confidence interval. The subjective nature of the test must also be taken into consideration when examining scores. The manual does not provide information regarding the construction or scoring of the CARS2-QPC, therefore, it should only be used to provide a clinician with more information on which to base the CARS2. The CARS2 should not be used as a sole source of diagnosis, but rather act as a piece of the assessment process.

References

- Dickerson, S., Calhoun, S. L., Murray, M. J., Morrow, J. D., Yurich, K. K. L., Mahr, F., Cothren, S., Purichia, H., Boudier, J. N., & Petersen, C. (2009). Comparison of scores on the checklist for autism spectrum disorder, childhood autism rating scale, and Gilliam Asperger's disorder scale for children with low functioning autism, high functioning autism, Asperger's disorder, ADHD, and typical development. *Autism & Developmental Disorders, 39*, 1682-1693.
- Eaves, R. C., & Milner, B. (1993). The criterion-related validity of the Childhood Autism Rating Scale and the Autism Behavior Checklist. *Journal Of Abnormal Child Psychology, 21*(5), 481-491. doi:10.1007/BF00916315
- Perry, A., Condillac, R. A., Freeman, N. L., Dunn-Geier, J., & Belair, J. (2005). Multi-site study of the Childhood Autism Rating Scale (CARS) in five clinical groups of young children. *Journal of Autism and Developmental Disorders, 35*, 625-634. doi: 10.1007/s10803-005-0006-9
- Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010) Childhood Autism Rating Scale, Second Edition (CARS2). *Western Psychological Services*.