

HTL 553

Reviews Of The Childhood Autism Rating Scale

Prepared by
Jeff Romanczuk

for
Dr. Bell

June 28, 2000

Reviews of the Childhood Autism Rating Scale

Background on and Description of Scale

The Childhood Autism Rating Scale was developed to fill the need for an observation instrument that could reliably distinguish children with autism from those with other developmental disabilities and also differentiate among the levels of severity along the autistic spectrum (Prizant, 1989). It was constructed by Schopler and Reichler in 1971 as a screening instrument to aid in diagnosing autism (Martin, 1989). In the 1970s, it was used only by its creators at North Carolina's TEACCH program for the Treatment and Education of Autistic and Related Communication Handicapped Children (Martin, 1989). The instrument became available to others in 1980 and the current version was first published in 1988. According to the online *Mental Measurements Yearbook*, it is \$25 for a manual and 25 instruments packaged together, \$8.70 for 25 scales and \$17.50 for the manual if purchased separately.

The CARS is based on five systems for diagnosing autism: Kanner's method (1943), Creak's (1961), and Rutter's (1978). It also uses the traits of autism described in the *Diagnostic and Statistical Manual of Mental Disorders* (the third edition, published in 1987, *DSM-III-R*) and the definition provided by the then National Society for Autistic Children (now the Autism Society of America). The manual accompanying the CARS describes how each of the fifteen items that make up the scale reflect the five sources used (Martin, 1989). Morgan (1988) adds that these scales account for age and are based on "consensual criteria" for autism gleaned from these sources. Although it was designed as a screening instrument and differentiation tool, the CARS can be used for treatment or education planning. See the attached (completed) CARS for the fifteen categories covered and the gradations of each.

Examinee Appropriateness

Prizant (1989) states that the CARS ratings are based on the child's response to various structured activities and situations, using a rating of 1 (for age appropriate) to 4 (severely abnormal for chronological age). The *DSM-III-R* (now the *DSM-IV*, 1994) provides detailed criteria for assessing Autism and differentiating it from other childhood onset disorders. However, the only quantification supplied is in the number of traits described that the patient exhibits. Since autism is diagnosed subjectively, the CARS helps to quantify this subjectivity (Prizant, 1989). Half point ratings can be given for each item, resulting in a range of 7 scores for each item and a total range of 15 (nonautistic) to 60 (severely autistic). Morgan (1988) adds that the number of 3 or above scores a child receives is also factored into this total, but is the only source to say this and does not describe how this tally of high ratings is used.

Usability, Ethical Propriety, and Teaching Feedback

The CARS can be used during or after direct observation, or created from a review of the child's records (Prizant, 1989). It is immediately usable with minimal training necessary for the rater, although she or he should have some experience with autism (Prizant, 1989). The results of the scale have been successfully applied to both appropriate treatment program initiation and educational planning. Since it is a screening instrument (Martin, 1989), it should not be the exclusive tool used in either case, but be part of a more thorough assessment of the patient or student. Another usability caution is that raters, especially if they are parents or teachers, should guard against having recent incidents influence the rating given for a particular item. If the "Observations" space at the bottom of each item is used correctly, though, it should keep the focus specific enough to be useful but broad enough to be used.

Concerning ethical propriety, the need to protect the information given applies to completed CARS scales even more so than to group-related or normative instruments. Anonymity cannot be given by talking in general trends for groups with the CARS. It is a very personal scale with information that should be protected at least as well as any individual student's school record. During classroom planning, only those legally allowed and with a valid need to know should see the completed

CARS. (That said, guarded dissemination of the attached completed CARS would be appreciated. JBR)

Soon after the most recent version of the CARS was published, Kurita, Miyake, and Katsuno (1989), created a Japanese translation of the instrument for use in Tokyo. Their literal Japanese translation was confirmed by a blind translation back to English. They used this instrument on 167 developmentally delayed children under sixteen years of age.

Test/Retest, Internal Consistency, and Interrater Reliability

Three kinds of reliability are described in the CARS manual: Test/retest (measuring the “temporal stability” of the results), internal consistency (the homogeneity across items when each could be worth various “credit” points), and interrater reliability (agreement in scale results among examiners of the same subject) (Prizant, 1989; Morgan, 1988). These data are based on Schopler’s own group of 280 subjects at TEACCH in the 1980s as assessed by two “trained” raters (Morgan, 1988).

The test/retest reliability for 91 of these cases assessed one year apart resulted in a correlation of .88. However, from the second to a third retest this coefficient kappa was .64. Prizant (1989) accounts for this by stating that the “developmental picture” for autism is still changing and improvement may be reflected in these results even though at each year’s retest, the subjects were being compared to peers their same chronological age.

The internal consistency (Cronbach’s coefficient alpha) given by Schopler is .94. The interrater reliability for two raters of all 280 subjects is .71 (Prizant, 1989). The most closely correlated item was “I. Relating to People” (at .93). The weakest correlation for the two raters was .55 for item “XIV. Level and Consistency of Intellectual Response” (Prizant, 1989).

For the first study using the Tokyo version of the CARS, the Cronbach alpha was slightly lower, but still very reliable at .87 (Kurita, Miyake, & Katsuno, 1989). Their interrater reliability ranged from .43 to .77. Although no mean was given for this range, it appears lower than Schopler’s mean of .71. Also, the total number of

subjects further confounds direct comparison; it is based on 167 subjects and two raters here and 280 subject and two raters in Schopler's case. The correlation likely was stronger than the numbers indicate for the Tokyo study, however, since one of the two raters showed a marked reluctance to use the half-point assessments, effectively making it a four-point rather than a seven-point scale (Kurita, Miyake, & Katsuno, 1989).

Sturmey, Matson, and Sevin (1992) compare the utility of the CARS to two other autism rating scales on 34 children diagnosed with Pervasive Developmental Disorders (the umbrella term from the *DSM-III-R* and *DSM-IV* that covers autistic spectrum disorder and four other diagnoses). Their Cronbach alpha coefficient was .851, but the inter-item correlation mean was only .283.

Content, Construct, and Concurrent Validity

According to Morgan (1988), there are three kinds of validity the CARS has been measured for: content (the degree to which the items are representative of the broader domain of behaviors sampled), construct (the extent to which the scale measures the psychological or theoretical construct under investigation), and concurrent (how well the instrument correlates with already available criteria on the subject: independent assessments done by child psychologists, the subjects other records, and parent interviews). In Prizant's (1989) view, content and construct validity are implicit in the consensual selection of items from five known definitions of autism. Administration of the Tokyo version saw a weak negative correlation between total score and age, even considering that subjects are compared only to their neurologically typical peers the same chronological age (Kurita, Miyake, & Katsuno, 1989). This is true despite Prizant's (1989) indication that appropriateness of behavior is harder to gauge among preschoolers than older children. However, the CARS is more like a behavior test than a content-based one. So the instrument should work better with younger children. Also, in each case the age comparison is based on the rater's experience or hunches, not on any prescribed normative traits. Even so, it is truly a "Childhood" Autism Rating Scale, not intended for use with

adults. The social, physical, and mental processes assessed have to be still in formation, not already set.

As for concurrent validity, Prizant (1989) notes that there is no significant difference in the correlation coefficient across various settings (unstructured interview, parent interview, psychoeducational testing, classroom observation, and records checking). The range of coefficient kappas for these was .63 to .84, with a mean concurrent validity coefficient of .80. Morgan (1988) adds discriminant validity (the ability of a scale to differentiate among various defined diagnostic groups) as a more specific subgroup of concurrent validity. Morgan (1988) notes that the CARS showed a hundred percent predictive accuracy in distinguishing between “trainable” (that is, mildly) mentally retarded and autistic. However, the CARS’ ability to discriminate among autism and the other pervasive developmental disorders was not as clear (Morgan, 1988). Martin (1989) notes that the manual accompanying the CARS says only that concurrent validity correlations are “high,” with very little detail on the research used to satisfy methodological questions.

Sturmey, Matson, and Sevin (1992) point out that especially with small experimental groups—only 34 children were in their sample—any item may decrease internal consistency while improving the predictive validity of the instrument. That is, an item can have important diagnostic significance yet lower the coefficient alpha. The CARS has perhaps only one such “rogue” item, “XIV. Intellectual Functioning.” But even this trait has demonstrated homogeneity with the other fourteen in most studies using the CARS.

Weaknesses (No Norming, Poor Explanations in the Manual, Signs of Age)

Most of the CARS’ shortcomings stem from the omissions rather than the commissions of the developers. Prizant (1989) finds fault with the scales having equal weight despite some items clearly not being as vital as others: “IX. Taste, Smell, and Touch Response and Use” vs. “I. Relating to People,” for example. Martin (1989) finds the manual lacking in several ways: the sparse concurrent validity detail mentioned above, how/why a score of 30 is the dividing line for autistic or non-autistic, and the

lack of normative data/tables given. Martin does acknowledge that the 1988 manual at least bases its information on 1,500 cases (not merely the original 280) and admits that normative data on “low incidence” handicapping conditions is always difficult to extrapolate. Even so, Martin believes that the manual’s narrative could say more about how the items relate to each other and the total score. Are the many traits of autism distinct or intermingled? Martin’s (1989) point is that the alpha coefficient should not be a substitute for a full explanation of internal consistency and the manual is the place to do this explaining. Another of the accompanying manual’s omissions stems from it being twelve years old. While the age is not reflected so much in the items themselves, it is obvious in the references. The National Society for Autistic Children has been the Autism Society of America since 1989. The use of the *DSM-IIIR* when the *DSM-IV* was published over six years ago is more noticeable. Even so, the described traits of autism did not change much in the 1994 version over the 1987 one. So the CARS as written remains usable and useful. What did change is the *DSM-IV*’s autism categorization among the pervasive developmental disorders and the level of detail given about autism.

Prizant (1989) points out a specific example of one of the manual’s few sins of commission. “For direct observation of the child’s response to pain, it may be necessary to pinch the child.” To Prizant’s criticism that there are surely other ways to get this information should be added that someone that unfamiliar with the subject should not be completing the CARS.

Strengths (“Demonstrated Psychometric Properties”)

Although the manual’s supporting detail is spare (Martin, 1989), Prizant (1989) notes that the instrument itself is better at discriminating cognitive impairment than it is for social impairments. Morgan (1988) adds that diagnosing autism has always been tough because its manifestations overlap with traits indicative of other disorders. Clinicians may fail to identify autism as such, or call something else autism. Although the latter error is more common in recent years, either mistake misleads parents and clients and confounds research. Even so, Morgan (1988) declared the CARS to be the strongest of the five autism scales he reviewed in terms

of its “demonstrated psychometric properties.” Morgan deems the CARS the front-runner due to its overall effectiveness. It uses a large sample, has explicit scoring criteria that account for age differences, has adequate content and construct validity, and has a high concurrent validity with independent clinical ratings.

References

- Kurita, H., Miyake, Y., & Katsuno, K. (1989). Reliability and validity of the childhood autism rating scale - Tokyo version (CARS-TV). *Journal of Autism and Developmental Disorders*, 19, 389-396.
- Martin, R. P. (1989). Review: childhood autism rating scale. *Journal of Psychoeducational Assessment*, 7, 368-369.
- Morgan, S. (1988). Diagnostic assessment of autism: a review of objective scales. *Journal of Psychoeducational Assessment*, 6, 139-151.
- Prizant, B. M. (1989). Review of the childhood autism rating scale. *Mental Measurements Yearbook* [on-line database through UTK library].
- Schopler, E., Reichler, R. J., & Renner, B. R. (1988). *The Childhood Autism Rating Scale*. Los Angeles: Western Psychological Services.
- Sturmey, P., Matson, J. L., & Sevin, J. A. (1992). Brief report: analysis of the internal consistency of three autism scales. *Journal of Autism and Developmental Disorders*, 22, 321-328.