

Validity and Reliability Analysis of Turkish Version of Childhood Autism Rating Scale



Seçil İNCEKAŞ GASSALOĞLU¹, Burak BAYKARA², Sibelnur AVCİL³, Yücel DEMİRAL⁴

SUMMARY

Objective: The purpose of this study is to expand validity and reliability analysis of Childhood Autism Rating Scale – Turkish Form (CARS-TF), whose internal consistency, content validity and discriminant validity for a sample group of limited size were examined by Sucuoğlu et al.

Method: 96 children and adolescents aged between 4-18, (48 diagnosed with pervasive developmental disorder (PDD) and 48 diagnosed with mental disability (MD) and developmental delay based on DSM-IV-TR criteria) were included in the study. In order to consider the reliability analysis of the Turkish Form of the scale, Cronbach's alpha values as internal consistency indicator, and inter-rater reliability and test-retest reliability were calculated. Principal components analysis and Varimax rotation were used in order to determine factors. The scale was compared with the Autism Behavior Checklist and Clinical Global Impression-Severity of Illness. The most appropriate cut-off point was determined for CARS-TF by conducting ROC analysis.

Results: As a result of principal components analysis and Varimax rotation, one component factor was obtained. Correlations between CARS-TF and the other scales were statistically significant. The Cronbach's alpha value of total score of the scale was determined to be 0.95. Test-retest reliability ($r=0.98$, $p<0.01$), and inter-rater reliability ($r=0.98$, $p<0.01$) were determined for the total score of the scale. The cut-off point of the scale was 29.5.

Conclusion: All of these results support that the scale adapted into Turkish is a valid and reliable assessment instrument.

Keywords: Pervasive Development Disorder, Childhood Autism Rating Scale, validity, reliability

INTRODUCTION

Pervasive developmental disorders (PDD) are neuropsychiatric disorders that start in the first years of life and involve specific delays and deviations in social skills, verbal and non-verbal communication, and cognitive development. These disorders are observed as insufficiency in mutual social interaction and communication, stereotypic behaviors, a shallow set of interests and limited functionality (Volkmar et al., 2002). According to the DSM-IV-TR diagnostic system,

PDD includes the Autistic disorder (AD), Rett syndrome (RD), Childhood disintegrative disorder (CDD), Asperger's syndrome (AS), and Not otherwise specified pervasive developmental disorders (NOS-PDD) (American Psychiatric Association, 2000).

In recent years, there has been an increase in the level of awareness of both society and physicians regarding pervasive developmental disorders. In addition, families are now more likely to report have an increasing possibility of reporting

Received: 04.08.2014 - Accepted: 22.05.2015

¹MD, Bursa Dörtçelik Child Hospital, Child and Adolescent Psychiatry, Bursa, ²Assoc. Prof., Dokuz Eylül University Medical Faculty, Child and Adolescent Psychiatry, İzmir, ³Assist. Prof., Adnan Menderes University Medical Faculty, Child and Adolescent Psychiatry, Aydın, ⁴Prof., Dokuz Eylül University Medical Faculty, Public Health, İzmir, Turkey.

e-mail: snuravcil@yahoo.com.tr

doi: 10.5080/u11197

their anxiety about autism in their children. Early diagnosis and intervention in PDD is very important in terms of etiological evaluation and guidance. Thus, it is important to determine the most convenient method for the early evaluation of autism. The suggested practice is to conduct a two-stage screening to examine the developmental problems in the first stage and autism-specific clinical features in the second stage (Johnson and Myers, 2007).

Numerous diagnostic instruments are used in evaluating pervasive developmental disorders. The diagnosis may be made by obtaining a detailed developmental history, directly observing the behavior and applying psychometric measurements specific to autism. Diagnostic instruments for autism are generally based on two main information sources: The first one involves information obtained from the caregivers about existing behaviors along with the developmental history, and the other one involves the direct observation of the behavior. The consistency of the data pertaining to diagnostic instruments is very important for both clinicians and researchers (Saemundsen et al., 2003).

Convenient screening scales to provide sensitive and specific information are critical as they pave the way for early diagnosis (Volkmar et al., 2002, Volkmar and Klin 2005). Childhood Autism Rating Scale (CARS), Autism Behavior Checklist and Social Communication Questionnaire (SCQ) are the scales that could be used in determining the severity of symptoms in screenings (Volkmar and Klin 2005). In Turkey, the Autism Behavior Checklist (Yılmaz Irmak et al., 2007) and Social Communication Questionnaire (Öner et al., 2012, Avcil et al., 2014) have been standardized. The Childhood Autism Rating Scale was translated and retranslated by Sucuoğlu et al., (1996) and was adapted into Turkish. This study investigated the internal consistency and item analysis of CARS-TF in the limited number sample group.

CARS was developed by Schoppler and Reichler in 1971. The 1988 edition of CARS has been used in this study, which is the 11th edition of the scale. The scale was finalized by evaluating more than 1500 cases over a course of more than 15 years. The first scale was developed as a research tool by the Child Research Project at North Carolina University based on the diagnostic criteria published by the British Working Party (Creak criterion). This first scale was named the Childhood Psychosis Rating Scale (CPRS). While developing CPRS, the goal was to minimize the diagnostic confusion concerning the definition of classic autism suggested by Kanner (1943). CPRS was rearranged to be used in the Treatment and Education of Autistic and Related Communication handicapped CHildren

(TEACCH) program in North Carolina and started to be called as CARS at the present time (Schoppler et al., 2007).

CARS is a 15-point behavioral rating scale developed to separate non-autistic children with mental retardation (MR) from children with autistic symptoms. It is especially effective on separating autistic children from children with trainable MR. It also enables clinicians to determine the severity of autism as mild-moderate and moderate-serious (Schoppler et al., 2007, Hergüner and Özbaran, 2010).

CARS should not be used as the final point. It should be used as the first step in understanding the characteristics of cases, determining their specific problems, and directing their personal treatments.

During the observation, the behaviors of child should be compared with a child who is of the same age and is developmentally normal. Abnormality, frequency, duration, and intensity of the behavior should be evaluated. The objective of this scale is to rate the behavior without etiological explanations. Some behaviors graded in CARS may not be specific to autism. Thus, to what extent the behavior deviates compared to normal should be evaluated without judging its reason (Schoppler and Reichler, 2007).

CARS emphasizes the observable data rather than the clinical intuition. CARS could be applied to children from all age groups as from the preschool period and forward. Rather than a subjective clinical judgement, it presents an objective and a measurable evaluation by directly observing the behaviors, which is among the important advantages of CARS (Teal and Wiebe, 1986).

Validity and reliability studies of CARS were conducted by Schoppler et al. within the scope of the TEACCH program between 1970 and 1980. The first clinical evaluation was applied to 537 cases. This application was performed during the first diagnostic interview of the Psychoeducational Profile (PEP) application. Watching the interviews made in one-way rooms with a mirror, the raters casted their scored immediately following the interview (Teal and Wiebe, 1986). The psychometric characteristics of CARS were firstly examined by Schoppler et al. (2007) between 1981 and 1983.

Translating and retranslating the scale in Turkey and obtaining its Turkish form, Sucuoğlu et al. (1996) evaluated the validity and reliability of the scale by using the methods of internal consistency and item analysis. The purpose of this study was to extend the validity and reliability analysis of the CARS-TF, which was examined by Sucuoğlu et al. (1996) in terms of internal consistency, content validity, and distinctiveness of the outlying groups of the sample group.

METHOD

Sample

The case group of the study consisted of children and adolescents aged 4-18 who were diagnosed with autistic disorder (n=37), not otherwise specified pervasive developmental disorder (n=8), Asperger's syndrome (n=1), Rett syndrome (n=1), and childhood disintegrative disorder (n=1) according to the diagnostic criteria of DSM-IV-TR.

The control group of the study involved children and adolescents aged 4-18 who were either diagnosed with mental retardation according to the diagnostic criteria of DSM-IV-TR or had a retarded level of general development in terms of age at the rate of 30% in the evaluation that was performed via the Ankara Developmental Screening Inventory (ADSI) and were observed to have no PDD diagnosis and symptoms according to the diagnostic criteria of DSM-IV-TR (n=48).

As three control cases were younger than 6 years of age with their developmental evaluation made via ADSI. Among 45 control cases older than 6 that had mental retardation, 33 had a mild MR (total intelligence quotient within the range of 50-70 in the WISC-R Intelligence Test), 8 had a moderate MR (total intelligence quotient within the range of 35-49 in the WISC-R Intelligence Test) and 4 had a severe MR (total intelligence quotient under 35 points in the WISC-R Intelligence Test).

The average age of the sample was 116.58 ± 45.27 months (9.70 ± 3.77 years). 20 participants were female (20.8%) and 76 were male (79.2%). There was no difference between the case and the control groups in terms of average age and gender distribution.

Data Collection Tools

Childhood Autism Rating Scale (CARS)

CARS is a 15-point behavioral rating scale developed to separate non-autistic children with mental retardation (MR) from children with autistic symptoms. It is especially effective on separating autistic children from children with trainable MR. It also enables us to determine the clinical severity of autism as mild-moderate and moderate-serious. Each item is graded with a half degree scoring between 1-4 (Schopler et al., 2007, Hergüner and Özbaran, 2010). The possible total score ranges from 15 at a minimum to 60 at maximum. According to the scoring, children with scores between 15 and 29.5 do not show autistic symptoms. Children with scores between 30-36.5 have a clinically mild-moderate autism and children between 37-60 have severe autism (Robert et al., 1988, Garfin

et al., 1988, Mesibov et al., 1989, Hergüner and Özbaran, 2010). In clinical cases where a score of 30 or more is obtained in those not diagnosed with AD according to DSM-IV-TR, it is required to evaluate the RD, CDD, AS, PDD-NOS diagnoses. AD patients obtaining a score below 30 are thought to have a mild-moderate disorder. It is suggested to use the cutoff scores differently in adolescents and adults. It is also suggested to use the limit as 28 points for autistic symptoms and 35 points for severe autistic symptoms (Schoppler et al., 2007, Hergüner and Özbaran, 2010).

CARS scoring may be made during the clinical interview or by means of intraclass observations, information obtained from parents, and from register records. The scoring should not be made without collecting all the required data. The person conducting the scoring should be informed about all the items and rating rules (Schoppler et al., 2007, Hergüner and Özbaran, 2010).

Autism Behavior Checklist

The Autism Behavior Checklist (ABC) is an assessment instrument involving 57 items and five subscales including sensorial area, communicating, use of body and objects, language skills, social and self-care skills. While the lowest score to be obtained from the scale is 0, the highest score is 159 (Krug et al., 1993). Yılmaz Irmak et al. (2007) translated the ABC into Turkish and determined that it was a valid and reliable instrument for Turkey. The cutoff score of the Turkish form of the scale was set at 39.

Clinical Global Impression-Disease Severity Scale

The Clinical Global Impression-Disease Severity Scale is a observer-scored tool developed in order to evaluate patients in clinical trials and observe the changes caused by the treatment in the process of follow-up. Clinical Global Impression (CGI) involves 3 parts including disease severity, recovery and side effect severity (Guy, 1976). In this study, only the disease severity section was used (severity index-SI).

Sociodemographic Data Form

The Sociodemographic data form involves data about the child such as the date of birth, whether she/he attends school or not, school success, peer relations, family rank among siblings, birth history, and developmental history. Regarding the family, it involves questions about the educational level and professional status of parents, number of children in family and history of physical or mental disorders of the parents.

Data Collection

Diagnostic evaluation of cases were conducted by a clinician that was experienced in mental retardation and autism according to the criteria of DSM-IV-TR. The Sociodemographic data form was completed for all cases. In order to measure the reliability among the raters, cases with PDD ($n=46$) who were in the main study sample and constituted the case group were rated by the raters (Seçil İncekaş Gassaloğlu; S.İ.G) and (Burak Baykara; B.B) on the basis of video records, completely blinded to each other. As two cases in the case group did not have a quality video record, these two cases were excluded from the study portion evaluating the reliability between the raters. Mothers of all cases were asked to read and fill the ABC themselves. After filling the scale, we asked if mothers “had questions pertaining to the forms”.

The clinician completed the CGI-SI form. In order to evaluate the test-retest reliability, CARS-TF was reapplied by the same observer (S.İ.G) to all the cases with PDD that were in the main study sample and constituted the case group ($n=48$) 60 (± 10) days after the first scale application.

Statistical Methods

The data were statistically evaluated by using SPSS Windows 15.0 packaged software and the statistically significant p value was determined as $p<0.05$.

In order to evaluate the reliability of CARS-TF, the Cronbach alpha values, item-total score correlations, reliability between the raters and the test-retest reliability were calculated as an indicator of the internal consistency.

In order to determine the factor structure of CARS-TF, the principal component analysis was applied to the data, according to the Kaiser normalization and Varimax transformation. In order to see the degree of the correlation between the items of CARS-TF, their correlations to one another were calculated.

In order to evaluate the criterion validity, the CGI-SI, CARS-TF scores evaluated in the same interview and the ABC scores filled by mothers in the case and control groups were compared by calculating the Pearson correlation coefficient.

ROC analysis was performed to determine the most convenient cutoff score of CARS-TF and additionally calculate the values like sensitivity, specificity, negative and positive predictive values.

In order to determine the content validity of the scale, the total CARS-TF scores of the case and control groups were compared by using the independent samples t -test.

RESULTS

Genders were matched in the case and control groups. Thus, the gender distribution showed no difference between the case group and the control group. While the case group involved 10 girls (20.8%) and 38 boys (79.2%), the control group involved 10 girls (20.8%) and 38 boys (79.2%). In addition, there was no significant difference between the age distributions of groups (Case group: 116.58 ± 45.27 months, control: 114.33 ± 43.24 months, $t=0.249$, $p=0.804$, independent samples t -test).

Reliability Analyses

In this study, the reliability measurements were performed using the internal consistency, reliability among the raters, and the test-retest measurements. Cronbach's alpha value was calculated to examine the internal consistency of CARS-TF. Being calculated separately for both raters, the Cronbach alpha coefficient was determined as 0.95 for both. In order to see the degree of the relationship between the items of CARS-TF, their correlations to each other were calculated (Table 1). Items of CARS-TF showed a good level of correlation except for item 14 (mental reaction level). Except for the item 14, the correlation coefficient values were calculated between 0.49 and 0.90, and all of them were determined to be statistically significant. Among the items of CARS-TF, the items of general impressions and human relations had the highest correlation ($r=0.90$), whereas item 14 showed an inverse correlation with other items. The items notifying the contribution of items to the scale and the total score correlation was 0.17 for item 14 and varied between 0.80 and 0.97 for the other items (Table 2).

The Pearson correlation coefficient was calculated in order to evaluate the test-retest reliability for the total score of CARS-TF, which was determined to be 0.98 and statistically significant ($p<0.01$). The CARS-TF was reapplied by the researcher (S.İ.G) 60 (± 10) days after the first application, with an increase of approximately 1.5 points in the total mean scores. This increase was found to be statistically significant ($p=0.000$) (Table 3). This increase observed in the total score of CARS-TF was determined to be statistically significant in the subtest of human relations ($p=0.044$) and in the subtest of non-verbal communication ($p=0.044$), which are among the subtests of the scale (Table 4).

In order to determine the reliability between the raters, the difference between the total mean scores of the first (S.İ.G) and the second (B.B) observers in CARS-TF and in the subtest mean scores, as well as the correlation coefficient between

Table 1. Correlation of CARS-TF items with each other

	1. Relationships with people	2. Imitation	3. Emotional reactions	4. Use of body	5. Object usage	6. Adapt to change	7. Visual response	8. Listen response	9. Taste, touch, smell response and usage	10. Fear or nervousness	11. Verbal communication	12. Nonverbal communication	13. Activity Level	14. The level of mental response	15. Overall impression
1. Relationships with people	1.00														
2. Imitation	0.86 *	1.00													
3. Emotional reactions	0.78 *	0.79 *	1.00												
4. Use of body	0.53 *	0.62 *	0.56 *	1.00											
5. Object usage	0.72 *	0.73 *	0.73 *	0.57 *	1.00										
6. Adapt to change	0.85 *	0.85 *	0.72 *	0.52 *	0.71 *	1.00									
7. Visual response	0.84 *	0.84 *	0.79 *	0.63 *	0.74 *	0.76 *	1.00								
8. Listen response	0.78 *	0.83 *	0.74 *	0.61 *	0.68 *	0.76 *	0.83 *	1.00							
9. Taste, touch, smell response and usage	0.60 *	0.61 *	0.49 *	0.57 *	0.67 *	0.58 *	0.65 *	0.60 *	1.00						
10. Fear or nervousness	0.79 *	0.80 *	0.71 *	0.51 *	0.63 *	0.83 *	0.72 *	0.65 *	0.53 *	1.00					
11. Verbal communication	0.88 *	0.87 *	0.79 *	0.59 *	0.72 *	0.79 *	0.84 *	0.76 *	0.60 *	0.78 *	1.00				
12. Nonverbal communication	0.85 *	0.86 *	0.82 *	0.55 *	0.75 *	0.77 *	0.86 *	0.76 *	0.64 *	0.73 *	0.84 *	1.00			
13. Activity level	0.61 *	0.71 *	0.68 *	0.56 *	0.56 *	0.65 *	0.67 *	0.71 *	0.49 *	0.64 *	0.61 *	0.66 *	1.00		
14. The level of mental response	-0.34 **	-0.40 *	-0.30 **	-0.37 **	-0.38 **	-0.42 *	-0.33 **	-0.30 **	-0.35 **	-0.40 *	-0.49 *	-0.31 **	-0.32 **	1.00	
15. Overall impression	0.90 *	0.87 *	0.82 *	0.63 *	0.73 *	0.78 *	0.80 *	0.76 *	0.64 *	0.75 *	0.86 *	0.86 *	0.65 *	-0.34 **	1.00

* $p < 0.01$, ** $p < 0.005$

Table 2. CARS-TF factor loadings and correlations between items and total scores

CARS Items	Single Component	Correlations between items and total score
Relationships with people	0.92	0.962
Imitation	0.94	0.963
Emotional reactions	0.87	0.926
Use of body	0.69	0.898
Object usage	0.82	0.933
Adapt to change	0.88	0.942
Visual response	0.91	0.970
Listen response	0.87	0.947
Taste, touch, smell response and usage	0.71	0.822
Fear or nervousness	0.84	0.804
Verbal communication	0.92	0.914
Nonverbal communication	0.91	0.965
Activity Level	0.76	0.796
The level of mental response	-0.45	0.178
Overall impression	0.92	0.971

Table 3. The mean total score of CARS-TF and total scores correlations were determined in the first and second interviews in case groups. The mean total score of CARS-TF and total scores correlations were determined by first and second observers in case groups

	The mean total score of CARS-TF* and SD	t**	p	Pearson correlation coefficient	p
1. interview	39.4±9.4	-6.661	0.000	0.98	<0.01
2. interview	40.9±10.1				
1. observers	40.71±10.31	0.804	0.426	0.97	<0.01
2. observers	40.42±10.96				

*Childhood Autism Rating Scale- Turkish Form

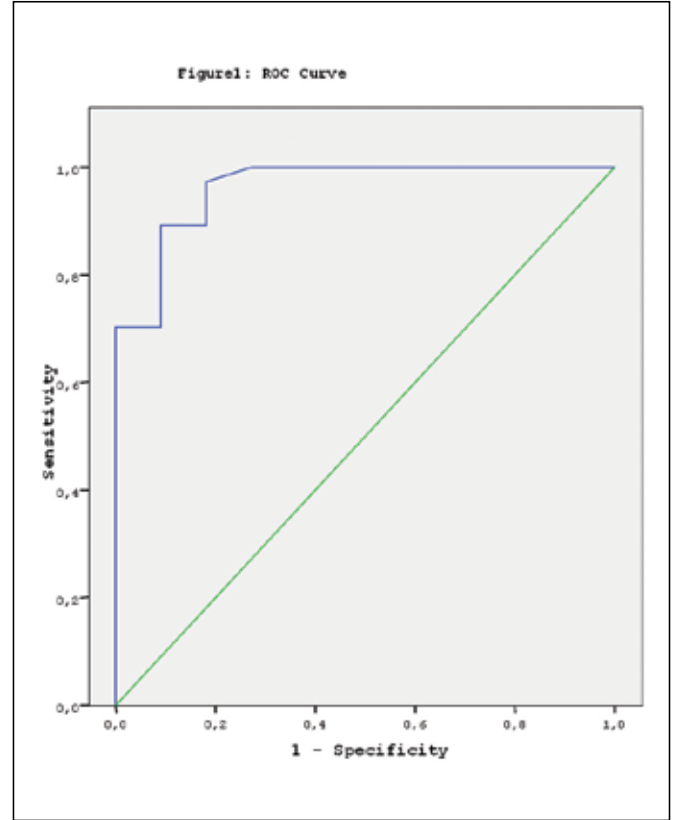
**For paired groups t-test

the raters were calculated and are present in Tables 3 and 4, respectively.

Validity Analyses

In order to determine the factor structure of CARS-TF, the principal component analysis was applied to the data according to the Kaiser normalization and Varimax transformation. It was found that it consisted of a single factorial structure except for item 14, which accounted for 70.79% of the total variance (Table 2).

It was determined that there was a highly significant correlation between the total score of CARS-TF and the total score



of CGI-SI ($r=0.87$, $p<0.001$) and a moderately significant correlation between the total score of the scale and the total score of ABC ($r=0.57$, $p<0.01$).

ROC analysis was performed in order to determine the cutoff score to be used to discern autism from MR. Figure 1 presents the results of the evaluations performed by using the analyses of the ROC curve. The reference line in the ROC curves shows the diagram obtained as a result of the incidental separation of both groups. In the diagram obtained by the values of sensitivity and 1-Specificity (wrong positiveness), the abundance of the area under the curve enables the test to create a better separation between the two groups. The area under the curve was determined as 96% (confidence interval 0.90-1.02; $p<0.0001$) for the total CARS-TF. The most convenient sensitivity (0.97) and specificity (0.82) values for the total CARS-TF were determined at a 29.5 cutoff score. While diagnosing the AD, it was determined that the sensitivity of the scale was 100% and the specificity was 81.8%. Considering the entire PDD group, it was found that the positive prediction value was 0.95 and the negative prediction value was 1 for the AD diagnosis. For all the PDD diagnoses, the sensitivity of the scale was 81.3% and the specificity was 100%. When evaluating the case and the control groups together, the positive prediction value was 1 and the negative prediction value was 0.84 for PDD.

Table 4. The difference of the subtest mean scores of CARS-TF*** and subtests of CARS-TF*** correlations were determined in the first and the second interviews in case groups. The difference of the subtest mean scores of CARS-TF*** and subtests of CARS-TF*** correlations were determined by first and second observers in case groups

	The difference of CARS-TF*** items scores in the first and second interviews				The difference of CARS-TF*** items scores of the first and second observers			
	The difference of the subtest mean scores and SD (n=48)	t**	p	Pearson Correlation Coefficient	The difference of the subtest mean scores and SD (n=46)	t**	p	Pearson Correlation Coefficient
Relationships with people	0.083±0.27	2.067	0.044	0.94*	-0.065±0.29	-1.521	0.135	0.94*
Imitation	0.0±0.17	0.00	1.00	0.98*	0.02±0.39	0.374	0.710	0.94*
Emotional reactions	0.05±0.38	0.927	0.359	0.88*	-0.01±0.34	-0.216	0.830	0.92*
Use of body	0.08±0.37	1.533	0.132	0.88*	0.07±0.40	1.265	0.212	0.87*
Object usage	0.03±0.26	0.829	0.411	0.95*	-0.11±0.44	-1.804	0.078	0.88*
Adapt to change	-0.01±0.26	-0.275	0.785	0.96*	0.18±0.53	2.361	0.023	0.84*
Visual response	0.04±0.26	1.071	0.290	0.93*	0.04±0.49	0.599	0.552	0.85*
Listen response	0.05±0.31	1.151	0.256	0.93*	-0.03±0.45	-0.489	0.627	0.89*
Taste, touch, smell response and usage	0.04±0.30	0.942	0.351	0.93*	0.16±0.50	2.185	0.034	0.86*
Fear or nervousness	0.08±0.37	1.533	0.132	0.88*	0.14±0.47	2.000	0.052	0.84*
Verbal communication	0.02±0.22	0.628	0.533	0.97*	0.00±0.40	0.000	1.000	0.91*
Nonverbal communication	0.08±0.27	2.067	0.044	0.95*	-0.09±0.35	-1.847	0.071	0.92*
Activity Level	-0.02±0.32	-0.443	0.659	0.89*	-0.08±0.46	-1.273	0.209	0.81*
The level of mental response	0.02±0.144	1.00	0.322	0.94*	-0.03±0.28	-0.771	0.445	0.75*
Overall impression	-0.02±0.10	-1.430	0.159	0.99*	0.10±0.33	2.221	0.031	0.90*

*** Childhood Autism Rating Scale- Turkish Form

** For paired groups t-test

*p<0.01

DISCUSSION

There is a need for both reliable and valid instruments to evaluate pervasive developmental disorders. As a result of the common need for a cost-effective and easily administered instrument, the validity and reliability of CARS, which meets all these features, have been investigated in numerous countries (Schopler et al., 1980, Pereira et al., 2008, Tachimori et al., 2003, Novardin et al., 1998). The scale was translated and retranslated by Sucuoğlu et al., (1996) in Turkey and created the Turkish version of the form. This study investigated the internal consistency and item analysis of CARS-TF in a sample group of 23 people, but required the investigation of other psychometric features of the scale in a larger sample group. The study of Sucuoğlu et al., (1996) provides limited information about the reliability of the scale. In order to implement the common use of the scale in Turkey, it is required to investigate both the validity and the reliability in a larger

sample group. The primary objective of this study was to investigate the validity and reliability of the CARS-TF in a larger sample group, in terms of its further psychometric features.

The internal consistency of the scale was evaluated in order to determine whether or not the feature to be measured could be measured with the scale in question. A greater value of the internal consistency reliability coefficient signifies that the items used for the measurement measures a homogeneous structure. The CARS-TF was determined to be highly reliable. Previous studies also reported a high reliability for CARS (Sucuoğlu et al., 1996, Schopler et al., 1980, Pereira et al., 2008, Tachimori et al., 2003, Novardin et al., 1998), which could be associated with the fact that all items of CARS measure features specific to autism.

In this study, except for the item 14 (mental reaction level), all the items of CARS-TF showed a good correlation with each other. This result is compatible with the studies of Magyar

and Pandolfi (2007) and Sucuoğlu et al. (1996), as well. This condition could be associated with the semantic features of the item 14. The item 14 in CARS-TF has a structurally and semantically different feature from other items in terms of scoring. As the severity of autism increased, the scores of other items except for item 14 increased. However, it requires extraordinarily superior mental skills to obtain high scores on item 14. It requires an impairment in the clinical presentation to obtain high scores from all other items and a mental advantage to obtain high scores from the item 14. Thus, the reverse correlation of the item 14 with other items is interpreted as an expected result.

In their study evaluating 91 cases at approximately 1 year intervals to measure test-retest, Schopler et al. (2007) determined the correlation coefficient as 0.88 ($p < 0.01$) with no significant difference in the mean scores of CARS one year later. Contrasting our results, this study has a statistically significant increase of approximately 1.5 points that was determined in the total mean scores of CARS-TF, which was reapplied by the rater (S.I.G) 60 (± 10) days after the first application. This score increase was associated with the intensification of symptoms in cases within a period of two months.

The construct validity of the CARS-TF was performed according to the principal component analysis method. As a result of the Varimax rotation, there was a unifactorial structure explaining 70.79% of the distribution, except for item 14. As a result of the examination that was performed using the principal component analysis method in the study of Di Lalla and Rogers (1994), a three-component structure included impairment in the social function, negative emotional response and change in the sensorial response. This three-component structure explained 64% of the total variance. Social impairment in this structure was determined as the largest component explaining 52% of the total variance and covering 10 CARS-TF items.

According to the principal component analysis, the components found in the study by Stella et al. (1999) comprised 64% of the total variance. This study determined the components of social communication, emotional reaction, social orientation, cognitive and behavioral consistency and inappropriate sensorial exploration behavior.

In their study, Magyar and Pandolfi (2007) determined 4 components explained 41% of the total variance according to the principal component analysis method. These components were determined as social communication, social interaction, stereotyped behaviors and sensorial abnormalities and emotional regulation.

Examining the literature, the results regarding the factor analysis indicate a significant difference. However, the structures defined in all factor analysis studies show a consistency with DSM-IV (Magyar and Pandolfi, 2007). CARS is a scale formed before the DSM-IV. Studies involving factor analysis support the fact that the scale is a convenient instrument for diagnosing autism in both the clinical and social group, in accordance with the diagnostic criteria of DSM-IV-TR (Magyar and Pandolfi, 2007). The difference between the findings in studies examining the factor structure of CARS-TF could be explained by the sample size and diagnostic differences in these studies.

All the items of CARS measure features specific to autism. As the group with no PDD does not have these features, the inclusion of this group in the measurement may cause all the items to seem like a single factor that acts similarly. This condition may disallow us to examine the relationship between the autism-specific features and the different groups within the autism spectrum disorders. When the CARS items are applied only to the PDD group, more than one factor appears (Magyar and Pandolfi, 2007, Stella et al., 1999).

The case/scale item rating is recommended to be at minimum, a 10 for the factor analysis (Baydur and Eser, 2006; Magyar and Pandolfi, 2007, Stella et al., 1999). In this study, the rating was 3.2, which was relatively small. The number of cases in our study did not provide the ideal case/scale rate for evaluating factor analysis. Thus, the single component that was revealed in the factor analysis in this study should be evaluated with suspicion. This result of our study could be associated with the relative insufficiency of the sample. There is a need for a larger sample group in order to evaluate the factor structure of the CARS-TF.

One of the methods used in evaluating the validity of a scale is the criterion validity. Being a part of the criterion validity, the concurrent validity compares the correlation coefficient between the result of an assessment instrument being developed and another assessment instrument that was previously developed. Both have high validity and measures the same feature.

This study compared the mean scores of ABC that was filled by mothers in the CGI-SI and the case group evaluated in the same interview and the mean scores of CARS-TF in order to evaluate the criterion validity. Examining the relationship between the CARS-TF items and the CGI-SI score, it was determined that except for item 14, all items showed a significant correlation with CGI-SI ($p < 0.01$). This study was the first to compare the CARS-TF items with CGI-SI. The correlation between the CARS-TF items, the total score of CARS-TF and CGI-SI, except for item 14, may be interpreted that

the scale items and the total score efficiently measure clinical severity. No significant correlation was determined between item 14 evaluating the mental reaction level and the CGI-SI scores. Item 14 does not measure the disorder severity and acts independently from the rest of the scale. In their study, Pereira et al., (2008) compared the score of the GAF (Global Assessment of Functioning) scale evaluating the general functionality level in similar to CGI-SI and the total score of CARS. As the disorder severity increases, the functionality score obtained from the GAF scale decreases and thus, GAF and CARS act reversely and are expected to show a reverse correlation. In their study, Pereira et al. (2008) determined that $r=-0.75$ ($p<0.001$), which signified that the total score of CARS could efficiently measure the impairment in functionality. This finding was compatible with our study.

The moderate-level correlation between CARS-TF and ABC is associated with the use of various methods like different scorings such as binary or likert type and the clinician's observation or the caregiver's evaluation.

In this study, we determined 100% sensitivity and 81.8% specificity for CARS-TF according to the diagnosis of AD and 81.3% sensitivity and 100% specificity for the scale according to the diagnosis of PDD. These results were compatible with current literature. According to the obtained statistical results, CARS-TF was considered as a good screening instrument for AD. In addition, there may not be as a convenient diagnostic instrument than AD for AS, CDD, and NOS-PDD.

A ROC analysis was performed to determine the cutoff score of the CARS-TF. The area remaining under the curve was determined to be 96% for the total CARS-TF. In other words, the scale accurately classifies the PDD and MR group with a probability of 96%. In our study, the total score of CARS-TF separated the PDD and MR group very well, which is similar to the study of Perry et al. (2005). While determining the cutoff score, the sensitivity and specificity values were expected to be high. While the increase of sensitivity signifies the increase of accurate positive proportion, the increase of specificity signifies the decrease of wrong positive proportion. Sensitivity for the CARS screening instrument, aims to discern individuals with autism-specific features in society from those without the features. A high sensitivity will cause individuals, who have no autism but some autism-specific features, to be diagnosed with autism as well. Specificity, on the other hand, will enable us to discern only those with autism among individuals with autism-specific features in society. A high specificity enables us to separate individuals with some autism-specific features in society from those being diagnosed with this disorder. As CARS is a screening instrument, it is

important to determine individuals demonstrating autism-specific features, even if not at a diagnosable level. If a person is mistakenly diagnosed with autism, this condition will probably be undone in the oncoming process. However, if the diagnosis is missed in a person with autism, this will result in the delay of convenient and timely treatment.

As study sensitivity is considered more important than specificity, the value 29.5 signifying a higher sensitivity was accepted as the cutoff score. The sensitivity (0.97) and specificity (0.82) values of this cutoff score are thought to be too high for a screening instrument. The cutoff score of the CARS-TF obtained from this study is compatible with the cutoff score (30) obtained from the original study (Schopler et al., 1980, 2007).

Study Limitations

This study involves some limitations to be emphasized. First of all, the number of individuals in the case and the control group is not enough for the factor analysis of CARS-TF. In this scale involving 15 items, a greater number of individuals in the case and the control groups would increase the statistical power of results. In order to evaluate the factor analysis of the CARS-TF, a larger sample group is needed.

Secondly, even though NOS-PDD is a more frequently encountered clinical condition among the PDD group as compared to autism, it is not considered sufficient in diagnosing and screening the disorders like CARS-TF, AS and NOS-PDD, which is an important limitation of the scale.

CONCLUSION

Autism is a lifelong disorder that affects individuals and families, resulting in important losses of mental capacity. The prevalence of autism has shown an increase within the last 20 years. In Turkey, there is a need for standardized screening and diagnostic instruments to better evaluate autism. Our study aimed to fill an important deficiency in this area. CARS is a valid and reliable assessment instrument that is used for diagnosis and screening in a number of countries and can determine the severity. It is thought that CARS will be preferred in studies concerning autism as it is used in discerning autism, as well as planning and evaluating the education, and is administered easily and in a short time to evaluate the different areas of development. CARS-TF was investigated by Sucuoğlu et al. (1996) in terms of validity and reliability to evaluate further psychometric features in our study. Results supporting that CARS-TF was highly valid and reliable were obtained.

REFERENCES

- Amerikan Psikiyatri Birlięi (2000) Psikiyatride Hastalıkların Tanımlanması ve Sınıflandırılması Elkitabı, Yeniden Gözden Geçirilmiş Dördüncü Baskı (DSM-IV-TR) (Çev. ed.:E Köroğlu). Hekimler Yayın Birlięi, Ankara, 2007.
- Avcil S, Baykara B, Baydur H et al (2015) 4-18 Yaş Aralığındaki Otistik Bireylerde Sosyal İletişim Ölçeęi Türkçe Formunun Geçerlik ve Güvenilirlik Çalışması. *Türk Psikiyatri Derg* 26:56-64
- Baydur H, Eser E (2006) Uygulama: Yaşam kalitesi ölçeklerinin psikometrik çözümlenmesi. *Saęlıkta Birikim* 1:99-123.
- DiLalla DL, Rogers SJ (1994) Domains of the Childhood Autism Rating Scale: relevance for diagnosis and treatment. *J Autism Dev Disord* 24:115-28.
- Garfin DG, McCallon D, Cox R (1988) Validity and reliability of the childhood autism rating scale with autistic adolescents. *J Autism Dev Disord* 18:367-78.
- Guy W (1976) ECDEU Assessment Manual for Psychopharmacology, Revised US Dept Health, Education and Welfare publication (ADM), Rockville, National Institute of Mental Health s.76-338.
- Hergüner S, Özbaran B (2010) Çocukluk Otizmini Derecelendirme Ölçeęi. Çocuk ve Ergen Psikiyatrisinde Ölçütler ve Ölçekler: Yaygın Gelişimsel Bozukluklar Yıkıcı Davranım Bozuklukları s. 37-40.
- Johnson CP, Myers SM (2007) Identification and evaluation of children with autism spectrum disorders. *Pediatrics* 120:1183-215.
- Krug DA, Arick JR, Almond PA (1993) Autism Screening Instrument for Educational Planning, Second Edition, Pro-edInc. Austin, Texas.
- Magyar CI, Pandolfi V (2007) Factor structure evaluation of the childhood autism rating scale. *J Autism Dev Disord* 37:1787-94
- Mesibov G, Schopler E, Schaffer B et al (1989) Use of childhood autism rating scale with autistic adolescents and adults. *J Am Acad Child Adolesc Psychiatry* 28:538- 41.
- Nordin V, Gillberg C, Nydén A (1998) The Swedish version of the Childhood Autism Rating Scale in a clinical setting. *J Autism Dev Disord* 28:69-75.
- Öner P, Öner Ö, Çöp E et al (2012) Sosyal İletişim Ölçeęinin Okul Öncesi Çocuklardaki Geçerlik ve Güvenilirlięi. *Klinik Psikofarmakol Bulteni* 22:43-50.
- Pereira A, Riesgo RS, Wagner MB (2008) Childhood autism: translation and validation of the Childhood Autism Rating Scale for use in Brazil. *J Pediatr* 84:487-94.
- Perry A, Condillac RA, Freeman NL et al (2005) Multi-site study of the childhood rating scale (CARS) in five clinical groups of young children. *J Autism Dev Disord* 35:625-34.
- Robert J, Reichler RJ, Roehen Renner B et al (1988) Practice DVD on Using the CARS, Western Psychological Services
- Saemundsen E, Magnússon P, Smári J et al (2003) Autism Diagnostic Interview-Revised and the Childhood Autism Rating Scale: convergence and discrepancy in diagnosing autism. *J Autism Dev Disord* 33:319-28.
- Schopler E, Reichler RJ, DeVellis RF et al (1980) Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *J Autism Dev Disord* 10:91-103.
- Schopler E, Reichler RJ, Roehen Renner B (2007) The Childhood Autism Rating Scale (CARS), 11. Baskı, Western Psychological Services.
- Sucuoglu B, Öktem F, Akkök F et al (1996) Otistik Çocukların Deęerlendirilmesinde Kullanılan Ölçeklere İlişkin Bir Çalışma. *3P Dergisi* 4:116-21
- Stella J, Mundy P, Tuchman R (1999) Social and nonsocial factors in the Childhood Autism Rating Scale, *J Autism Dev Disord* 29:307-17.
- Tachimori H, Osada H, Kurita H (2003) Childhood autism rating scale-Tokyo version for screening pervasive developmental disorders. *Psychiatry Clin Neurosci* 57:113-8.
- Teal M, Wiebe MJ (1986) A Validity analysis of selected instruments used to ases autism. *J Autism Dev Disord* 16:485-94.
- Volkmar FR, Klin A (2005) Issues in the Classification of Autism and Related Conditions. *Handbook of Autism and Pervasive Developmental Disorders, Cilt 1, Diagnosis, Development, Neurobiology, and Behavior*, 3. Baskı, FR Volkmar, R Paul A Klin (Ed). John Wiley&Sons, Inc., s. 5-42.
- Volkmar FR, Lord C, Klin A et al (2002) Autism and the Pervasive Developmental Disorders. *Child and Adolescent Psychiatry, a Comprehensive Textbook*, M Lewis (ed), Baltimore. Lippincott Williams and Wilkins.
- Yılmaz İrmak T, Tekinsav Sütçü S, Aydın A et al (2007) Otizm Davranış Kontrol Listesinin (ABC) Geçerlik ve Güvenilirlięinin İncelenmesi. *Çocuk ve Gençlik Ruh Saęlığı Dergisi* 14:13-23.