

## Exploring the Validity and Reliability of Turkish Version of Gilliam Autism Rating Scale-2: Turkish Standardization Study\*

### Gilliam Otistik Bozukluk Derecelendirme Ölçeği-2 Türkçe Versiyonu'nun (GOBDÖ-2-TV) Geçerlik ve Güvenirliğinin Araştırılması: Türkiye Standardizasyon Çalışması

İbrahim H. DİKEN  
Anadolu University\*  
Özlem DİKEN  
Anadolu University\*\*\*

Avşar ARDIÇ  
Pamukkale University\*\*  
James E. GILLIAM  
Austin\*\*\*\*

#### Abstract

This study aims at exploring the psychometric characteristics of the Turkish version of the Gilliam Autism Rating Scale-2 (TV-GARS-2) in order to standardize it into Turkish. Individuals diagnosed with autism, intellectual disability, hearing impairment and typically developing children were the participants of this standardization study (n=1191). After carrying out the Turkish translation procedure, the reliability and validity of TV-GARS-2 were examined by conducting a series of analyses such as construct validity, discriminant validity, item analysis, confirmatory factor analysis, internal consistency and test-retest reliability. Results yielded that TV-GARS-2 is a reliable and valid assessment tool that can be used with individuals with autism in Turkey.

*Keywords:* Autism, assessment, GARS-2, validity, reliability, Turkish sample.

#### Öz

Bu çalışmada Gilliam Otistik Bozukluk Derecelendirme Ölçeği-2 Türkçe Versiyonu'nun (GOBDÖ-2-TV) Türkçe standardizasyonu için GOBDÖ-2-TV'nin psikometrik özellikleri araştırılmıştır. Çalışmada 1191 otistik bozukluk (otizm) tanılı bireyden veri toplanırken, ayırt edici geçerlik için ayrıca zihinsel, işitme yetirsizliği olan ve yetersizliği olmayan/normal gelişim gösteren bireylerinden de veri toplanmıştır. Türkçe çeviri ve dil eşitliği çalışmalarından sonra, yapı geçerliğinin sınanması, ayırt edici geçerliğin sınanması, doğrulayıcı faktör analizi, iç tutarlılık ve test-tekrar test güvenirliliği gibi bir seri analiz ile GOBDÖ-2-TV'nin geçerlik ve güvenirliliği sınanmıştır. Bulgular, GOBDÖ-2-TV'nin Türkiye'de otistik bozukluk (otizm) gösteren çocukların değerlendirme sürecinde kullanılabilecek yeterli geçerlik ve güvenirlilik özelliklerine sahip olduğunu göstermiştir.

*Anahtar Sözcükler:* Otistik bozukluk (otizm), değerlendirme, GARS-2, geçerlik, güvenirlilik, Türkiye örnekleme.

\* Ph.D., Dr. İbrahim H. DİKEN Anadolu University, Faculty of Education, Department of Special Education, Eskisehir, Turkey; ihdiken@anadolu.edu.tr

\*\* M.A., Ph.D. Student, Psychologist, Avşar ARDIÇ Pamukkale University, Faculty of Education, Department of Special Education, Denizli, Turkey; aardic@pamukkale.edu.tr

\*\*\* Ph.D., Özlem DİKEN Speech and Language Therapist Anadolu University, Faculty of Education, Department of Special Education, Eskisehir, Turkey; odogramaci@anadolu.edu.tr

\*\*\*\* Ph.D., James E. GILLIAM Retired Professor, Austin, TX, USA; jgilliam09@gmail.com

\*The authors would like to thank all teachers and parents who contributed the study by filling out the TV-GARS-2 forms for individuals with autism and Dr. Güçlü Şekercioğlu and Dr. Şener Büyüköztürk for their support and suggestions on the statistical analysis of the TV-GARS-2.

Correspondence concerning this article should be addressed to İbrahim H. Diken, Anadolu University, Faculty of Education, Department of Special Education, Eskisehir, Turkey, 26470. E-mail: ihdiken@anadolu.edu.tr

### Introduction

As a lifelong, trainable, and developmental disorder that shows its characteristics before the age of three, (Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision; *DSM-IV-TR*, 2000; Filipek et al., 1999, 2000; International Classification of Diseases-10; *ICD-10*, 1993; Volkmar et al., 1994), autism is a disorder classified among a group of disorders under Autism Spectrum Disorders (ASD) or Pervasive Developmental Disorder (PDD) (*DSM-IV-TR*, 2000). According to the Center for Disease Control and Prevention (2012), the prevalence of autism is 1 in 88 in the us. Limitations in (1) social interaction, (2) communication, and (3) restricted, repetitive, and stereotyped behavior, interest, and activity patterns are three common areas with which individuals with autism have problems (*DSM-IV-TR*, 2000).

Autism is known to occur around the world regardless of differences in race, culture, and economic class (Trembath, Balandin, & Rossi, 2005). In order to be eligible to receive special education or related treatment services, early identification or diagnosis process is crucial for children with special needs including autism. Therefore, screening and assessment procedures in terms of the child-find process have to happen as early as possible (Filipek et al., 1999; Stone & DiGeronimo, 2006; Volkmar et al., 1999). Assessment and treatment procedures in ASD have two main areas in which several professional groups and professionals have been interested. *American Academy of Pediatrics* (The Pediatrician's Role in the Diagnosis and Management of Autistic Spectrum Disorder in Children, 2001; Technical Report: The Pediatrician's Role in the Diagnosis and Management of Autistic Spectrum Disorder in Children, 2001; Identification and Evaluation of Children with Autism Spectrum Disorders: Guidelines for the Clinician Rendering Pediatric Care: Clinical Report by Johnson, Myers, & the Council on Children with Disabilities, 2007), *the Child Neurology Society and American Academy of Neurology* (The Screening and Diagnosis of Autistic Spectrum Disorders by Filipek et al., 1999, 2000), and *American Academy of Child and Adolescent Psychiatry* (The Practice Parameters for the Assessment and Treatment of Children, Adolescents, and Adults with Autism and Other Pervasive Developmental Disorders by Volkmar et al., 1999) have supported studies and provided valuable information regarding assessment processes for individuals with autism (cited in Al Jabery, 2008). Among these studies, the practice parameter study (Filipek et al., 1999, 2000) stated two levels for the screening and assessment process (level one and level two). The focus of the level one is the screening process whereas the purpose of level two is the process of diagnose. Johnson, Myers, & the Council on Children with Disabilities (2007) and Filipek et al., (1999, 2000) recommended specific screening tools regarding autism for professionals to use when they have concerns about the individual. Gilliam Autism Rating Scale (GARS) and Gilliam Autism Rating Scale-2 (GARS-2) as level two instruments have been recommended for at risk children who are 18 months or older. In addition, Coonrod and Stone (2005), and Lord and Corsello (2005) recommended Gilliam Autism Rating Scale (GARS) as non-age specific measure.

Gilliam Autism Rating Scale-2 (GARS-2, 2005), revised version of Gilliam Autism Rating Scale (1995), was recommended to be used as type -two (level two) assessment instruments by Johnson, Myer, and the Council on Children with Disabilities guidelines (2007). The GARS and GARS-2 have been used in several studies (e.g., Al Jabery, 2008; Hodge, 2008; Lecavalier, 2005; Mazefsky & Oswald, 2006; Phillips, 2009; Schreck & Mulic, 2000; South et al., 2002; Tafiadis, Loli, Tsanousa, & Tafiadi, 2008). In two of these studies, GARS-2 was adapted in two different cultures or languages. Al Jabery (2008), for example, in his study, examined the reliability and validity of the GARS-2 with Jordanian population in Jordan while Tafiadis, Loli, Tsanousa, and Tafiadi, (2008) studied the psychometric characteristics of the GARS-2 with Greek population in Greece. Results of both of the studies showed that adapted GARS-2 had high level of reliability and validity characteristics.

There are relatively few tests available outside of English speaking countries for identifying and assessing autism. The instruments that are available are often translations of tests that were developed and normed in English speaking countries (e.g., Al Jabery, 2008). There is currently no

standardized or norm-referenced assessment tool in Turkey to screen, diagnose or to use in the assessment practices of children and adults with autism although reliability and validity of some scales such as Childhood Autism Rating Scale-CARS (Sucuoğlu, Akkök, & Gökler, 1996; İncekaş, 2009), The Checklist for Autism in Toddlers-CHAT (Tetik Kabil, 2005), The Modified Checklist for Autism in Toddlers-M-CHAT (Yıkgeç, 2005; Ünal, et al., 2006), and Autism Behavior Checklist-ABC (Yılmaz-Irmak, Tekinsav-Sütçü, & Aydın, 2007) were studied with small Turkish samples. Therefore, there is pressing need for standardized or norm-referenced assessment tool(s) for the assessment of individuals with autism in Turkey. Hence, this study was designed to provide a valid and reliable level one or two assessment tool to be utilized by professionals to diagnose children and young adults with autism and to provide support for their families for accurate and appropriate assessment practices in Turkey. The purpose was to investigate psychometric characteristics of the Turkish Version of the Gilliam Autism Rating Scale-2 (TV-GARS-2) in order to standardize it into Turkish.

## Method

### *Participants*

Participants included 1191 children and young adults who were diagnosed with autism. Out of 1191, 929 (78%) of them were males and 262 (22 %) were females. Participants' age ranged from 3 to 23 years with a mean of 9.88 (SD=3.95). Participants were selected in a variety of ways. Parents, teachers, psychologists and other professionals working in state schools for individuals with autism called Independent Education Center for Individuals with Autism (OCEM) and private special education and rehabilitation centers in twenty-one cities of Turkey were contacted by the first author and asked to complete the GARS-2 for individuals who had a formal diagnose of autism. The following criteria were used for determining the research group in the study: (a) having been diagnosed of autism according to psychological testing done in Counseling and Guidance Centers and individuals' health reports, (b) being between ages of 3 and 23, (c) residing in Turkey. Applying these criteria, 1428 participants were reached. However, 237 participants were excluded from the study because of the missing items that were left by these participants. The final normative sample included 1191 children and young adults with autism. Out of 1191 forms, 706 (59.3%) forms were filled out by teachers, psychologists, or other specialists working with individuals with autism, 485 (40.7%) forms were filled out by family members of individuals with autism. During the study, for the purpose of examining the discriminant validity of the TV-GARS-2, data were also collected from children and young adults who did not have autism but had either been diagnosed with intellectual disability (n=39) or hearing impairment (n=50) and from children with normal development (n=44).

### *Measure*

*Gilliam Autism Rating Scale-2 (GARS-2, 2005)*. With 14 items in each and a total of 42 items, GARS-2 had three subscales called stereotyped behaviors, communication, and social interaction. Definition of autism (the Autism Society of America) and diagnostic criteria for the autism (DSM-IV-TR, 2000) were used while developing items for GARS-2. As indicated in its manual (Gilliam, 2005), GARS-2 can be used for the following five purposes: (a) identifying persons who have autism, (b) assessing persons referred for serious behavior problems, (c) documenting progress in the areas of disturbance as a consequence of special intervention programs, (d) targeting goals for change and intervention on a student's Individualized Education Plan (IEP), and (e) using in research projects with individuals with autism. Original GARS-2 was normed on a sample of 1,107 children and young adults aged between 3 and 22 who had been diagnosed as autism. In the original study of the GARS-2, reliability and validity were examined by carrying out a series of psychometric procedures such as content sampling and time sampling for reliability, content-description, criterion-related, and construct-identification validity. Results of these analyses

revealed that GARS-2 a psychometrically sound screening instrument (Gilliam, 2005).

*Translation procedures.* Before conducting the validity and reliability studies of the TV-GARS-2, translation procedures were carried out. Six professionals from a variety of related disciplines such as speech and language therapy, psychology, guidance and psychological counseling and special education took part in this process. Translations from English to Turkish and then back translations from Turkish-English-Turkish were carried out. Perceptions of professionals were gathered on Turkish items of the scale and final form of the scale was established. A sample of fifteen mothers who had a child with autism and had different levels of education filled the final form in order to check whether Turkish translations of the items were appropriate for raters who had different levels of education. After this process, items were finalized to be used in validity and reliability studies of the TV-GARS-2.

#### *Pilot study*

Before carrying out standardization study of the TV-GARS-2, a pilot study (Diken, Gilliam, Ardic, Diken, & Sweeney, 2012) was carried out. Participants of this study consisted of 436 individuals diagnosed with autism (331 male and 105 female, mean of ages was 8.01 with  $SD=3.77$ ), individuals diagnosed with intellectual disability ( $n=44$ ), with hearing impairment ( $n=49$ ), and typically developing individuals ( $n=44$ ). Validity and reliability of the TV-GARS-2 were explored by conducting a series of analyses except for confirmatory factor analysis (CFA) that was explored in the current study. Results yielded that TV-GARS-2 is a reliable and valid assessment tool to be further explored to be used with individuals with autism in Turkey (Diken, Gilliam, Ardic, Diken, & Sweeney, 2012).

## Results

#### *Validity of the TV-GARS-2*

*Content-description validity and item analysis.* Item-discrimination analysis was conducted to confirm the validity of the test items. Two item discrimination criteria were used to test the TV-GARS-2. The item-discrimination coefficients were determined as at or beyond .05 level by using the criteria established by Hammill, Brown, & Bryant (1992). Second, at least half of the correlation coefficients had to reach or exceed .35 in magnitude. The minimum is large enough to ensure that each item is making a meaningful contribution to the subtest. Conventional item analysis was performed on 925 cases from the sample. These cases were selected because they completed all 42 items of the TV-GARS-2. In most cases, item analyses are performed for each age interval, but because non-significant relationship exists between age and scores on the TV-GARS-2 subscales, item analyses were not necessary at each age. The results regarding item-discrimination coefficients are reported in Table 1. The following median coefficients were obtained; Stereotyped Behavior, .38; Communication, .28; Social Interaction, .48. The median coefficient with sum of all items was .38. The median coefficients for the subscales and the total were statistically significant ( $p<.01$ ). In addition, they were well beyond the minimum criteria for magnitude and provided ample evidence of content-description validity.

Table 1.

*Item-discrimination coefficients for the TV-GARS-2 items*

Stereotyped Behavior		Communication		Social Interaction	
Item #	r	Item #	r	Item #	r
1	.41	15	.28	29	.48
2	.12	16	.25	30	.45
3	.37	17	.27	31	.41
4	.32	18	.30	32	.39
5	.28	19	.36	33	.49
6	.35	20	.49	34	.47
7	.42	21	.33	35	.45
8	.40	22	.24	36	.56
9	.38	23	.23	37	.54
10	.38	24	.16	38	.48
11	.36	25	.22	39	.46
12	.39	26	.43	40	.49
13	.43	27	.30	41	.53
14	.36	28	.21	42	.23
Median = .38		Median = .28		Median = .48	

*Construct validity.* To demonstrate the construct validity of a test, one must delineate as fully as possible the variable (construct) that the test purports to measure. This is done by setting up hypotheses are subjected to scientific investigation, and they are accepted or rejected on the basis of the results. The following hypotheses, as tested in the original study of the GARS-2 in the US and in a pilot study (Diken, Gilliam, Ardic, Diken, & Sweeney, 2012), were tested for construct-identification validity of the TV-GARS-2:

1. Because the behaviors measured by the TV-GARS-2 reflect the lifelong nature of autism, TV-GARS-2 scores should not correlate highly with chronological age.
2. Because the TV-GARS-2 subscales are related to each other (i.e., they all contain items that measure some aspect of autism), the subscales of the TV-GARS-2 should be positively related to each other.
3. Because the items within each TV-GARS-2 subscale measure similar traits, the items of a subscale should relate highly with the total score of that subscale.
4. Because the TV-GARS-2 subscales all measure characteristics of autism, they should be positively related to the Autism Index.
5. Because the TV-GARS-2 measures autism, the scores of persons with autism should differ significantly from those of persons who do not have autism.
6. Because autism is known to occur around the world regardless of race, culture, and economic class, results of the structure of the GARS-2 will be similar with Turkish samples based on Confirmatory Factor Analysis (CFA).

*Hypothesis 1: Relationship of the TV-GARS-2 subscales to age.* To study this hypothesis, the TV-GARS-2 raw scores were correlated with age using data of all participants with autism. Coefficients were found .00 ( $p > .01$ ) for Stereotyped Behaviors, -.02 ( $p > .01$ ) for Communication, .02 ( $p > .01$ ) for Social Interaction, and .02 ( $p > .01$ ) for the total (sum of all 42 items).

*Hypothesis 2: Interrelationship among GARS-2 subscales.* To examine the relationships of the TV-GARS-2 subscales, standard scores of the subscales were correlated for all the participants. For the TV-GARS-2 subscales, the mean has been set at 10 and the standard deviation at 3 as in the original GARS-2. Standard scores for the TV-GARS-2 subscale are derived directly from a cumulative frequency table containing the raw scores received by the normative sample. When normative tables are constructed, the raw scores are transformed into the desired derived distribution (i.e., into a distribution with a mean of 10 and a standard deviation of 3).

Table 2 displays the results of the correlations of the TV-GARS-2 subscales. All of the

correlations are moderate to large in magnitude by Hopkins's criteria. Clearly, the items of each subscale measure the same construct (i.e., behavioral characteristics of autism).

Table 2.

*Correlation of TV-GARS-2 subscale standard scores (Decimals omitted)*

	Stereotyped Behaviors	Communication
Communication	24'	-
Social interaction	62'	36'

\*p < .01

*Hypothesis 3: Validity of TV- GARS-2 items.* The discriminating power of an item, computed by the point-biserial method of item-total correlations, is sometimes referred to as item validity because these coefficients reflect the degree to which the items of a subtest or test are measuring the same constructs. These data can be cited as evidence of a test's reliability because strong item discrimination can only result from strong construct validity. Evidence from the item validities associated with the TV-GARS-2 appear in Table 1. As can be seen from Table 1, almost all items meet the criteria accepted at the literature as acceptable item-total coefficients. Discussion of these results will be addressed at discussion section.

*Hypothesis 4: Relationship of subscale standard scores to the Autism Index.* To test this hypothesis, standard scores for all subscales of the participants in the sample were correlated with the Autism Index, another type of normalized standard score which has a mean of 100 and a standard deviation of 15 (as set in the original GARS-2) and represents the TV-GARS-2's overall assessment of the characteristics of autism manifested by an individual. Summing the standard scores for all subscales of the TV-GARS-2 that were recorded derives the Autism Index. Part-whole correlations were computed, meaning that when a subscale was correlated with the Autism Index, its values were excluded from the index before the correlation was computed. In this way, the true correlation of the subscale value to the Autism Index is verified. As can be seen in Table 3, all correlations are large in magnitude and significant (p < .01).

Table 3.

*Correlation of TV-GARS-2 subscales with the Autism Index (Decimals omitted)*

	Stereotype Behaviors	Communication	Social Interaction
Autism Index	.46'	.33'	.54'

\*p < .01

*Hypothesis 5: Differences on TV-GARS-2 standard scores between diagnostic groups.* In terms of the TV-GARS-2, individuals who will be tested with this instrument will be those with severe behavioral disorders and developmental disabilities. Therefore, to establish the validity of the TV-GARS-2, the TV-GARS-2 was administered to a sample of individuals without autism (N=133). These participants were diagnosed with intellectual disability (n= 39) and hearing impairment (n=50). A group with normal development (n= 44) was also selected to serve as a control group. To test the hypothesis that persons from different diagnostic groups will score differently than persons with autism on the TV-GARS-2, one-way analysis of variance procedure was used but homogeneity of variance (Levene Test) was not assumed. Therefore, appropriate procedures Dunnet C were run. Results showed that differences between groups were statistically significant. Table 4 reports the mean standard scores for all groups. On each TV-GARS-2 subscale and the Autism Index, the group with autism received significantly higher scores (p<.01) than

the other diagnostic groups. On the Autism Index, the group with normal development received significantly lower ratings than the other diagnostic groups.

Table 4.

*Mean of standard scores of different diagnostic groups on the TV-GARS-2*

Diagnostic Group	N	TV-GARS-2 Subscale Standard Scores			
		Stereotyped Behaviors	Communication	Social Interaction	Autism Index
Intellectual disability	39	5	5	4	65
Hearing impairment	50	4	4	4	61
Normal development	44	3	3	3	56
Autistic disorder	50	10	10	10	100

*Hypothesis 6: Results of CFA on the TV-GARS-2.* Out of 925 participants who completely filled out the TV-GARS-2 forms for individuals with autism, 200 were selected randomly for carrying out the Confirmatory Factor Analysis (CFA) for further exploring construct validity of the TV-GARS-2. Results indicated that relationships between latent and observed variables were mostly high. Results also yielded that degree of freedom (df) was found as 813 while chi-square value was 1730.08. The results of distribution of chi-square to df [chi-square/df(1730.08/813)] was found as 2.13. Having a value of distribution of chi-square to df under 2.5 is suggested as a perfect chi-square goodness-of-fit (Çokluk, Şekercioğlu, & Büyüköztürk, 2010; Kline, 2000). Results also revealed that Root Mean Square Error of Approximation (RMSEA) value was found to be .071 and Comparative Fit Index (CFI) was found to be .89. All these results confirmed the structure of TV-GARS-2.

#### *Reliability of the TV-GARS-2*

*Content Sampling.* The internal consistency reliability of the items on the TV-GARS-2 was investigated using the Cronbach's alpha coefficient. Coefficient alphas were computed for all of the subscales of the TV-GARS-2 using the data obtained from all of the participants. The resulting coefficients were .79 for Stereotyped Behaviors, .77 for Communication, .85 for Social Interaction, and .88 for the total test (all 42 items).

Table 5.

*Results of time sampling of the TV-GARS-2*

TV-GARS-2 Subscales	Time 1		Time 2		Correlations r
	M	SD	M	SD	
Stereotyped behaviors	10	3	10	3	.98
Communication	9	3	9	3	.99
Social interaction	10	3	9	3	.99
Autism Index	98	17	97	17	.99

$p < .01$ ,  $r^2$  = Coefficient corrected for restricted range

*Time Sampling (Test-retest reliability).* To determine whether the results of the TV-GARS-2 were stable over time, a study was completed in which raters completed the TV-GARS-2 twice, 3 weeks apart. A total number of 42 individuals with autism enrolled in a private special education and rehabilitation center in Eskisehir in Turkey took part in this procedure. The raters were parents of the children. The mean age of the children was 8 years (SD=2.98). Thirty-one of these children were males and eleven were females. Raw scores for the two testing were converted into standard scores and indexes. The values were then correlated and corrected for restriction in range. The results, reported in the Table 5, provide evidence of the stability of the TV-GARS-2 when used

with individuals with autism. The coefficients are all beyond the .01 level of significance and of sufficient magnitude to suggest that the TV-GARS-2 has good test-retest reliability for use as an instrument for identifying persons with autism. These findings demonstrate that the TV-GARS-2 yields results that are stable over time. The resulting coefficients were .98 for Stereotyped behaviors, .99 for Communication, .99 for Social interaction, and .99 for autism index.

### Discussion

In this study, we aimed at assessing the psychometric characteristics of the Turkish version of GARS-2 to be utilized by professionals in Turkey. For this aim, we studied the validity and reliability of the Turkish version of GARS-2. Participants in the normative sample covered a wide geographic range coming from 21 cities around Turkey. This diversity adds to the strength of the subscales and provides comparisons for a variety of relevant demographic characteristics. In the study, professionals ( $n=706$ ) completed the subscales more than parents ( $n=485$ ). Since teachers and other professionals will be the principal raters using the TV-GARS-2, this ratio approximately 1.46:1 appears suitable. Moreover, the original GARS was also found to be useful and recommended as an appropriate instrument for gathering information from parents about their children with autism in the US (Filipek et al., 2000). Since there is no information about the prevalent characteristics of individuals with autism in Turkey, we cannot decide whether our sample represent population with autism in Turkey. However, given the male-to-female ratio (3.15:1 in the US) as a reference, the sample used for this study appears to be representative.

According to Hammill et al., (1992), normative samples should have at least 75 to 100 participants at every age level. They also state that total sample for this kind of study should be at least 750 to 1000. The normative sample for the TV-GARS-2 meets the standard for total sample size and but did not meet the standard for the minimum of 75 participants at all of ages. Results of the current study revealed that there were minimal differences among participants at different age levels. However, the relationship was statistically not significant at neither .01 nor .05 significance levels. This was not surprising because the behaviors of autism are not known to differ in terms of age (American Psychiatric Association, 1994). Statistical analyses were undertaken to confirm these observations. Correlations of subscale raw scores with age resulted in correlations of .08 for stereotyped Behaviors, .15 for Communication ( $p<.01$ ), and .06 for Social Interaction. Although the correlation for the Communication subscale with age is significant, its magnitude is small according to Hopkins' Likert-scale approach in determining the magnitude of coefficients (As cited in Gilliam, 2005). He suggested that coefficients between 0.0 and .09 are very small, coefficients between .1 and .29 are small, coefficients between .3 and .49 are moderate, coefficients between .5 and .69 are large, and coefficients between .7 and .89 are very large. Using Hopkins's guidelines, one must conclude that the relationship between subscale scores and age on the TV-GARS-2 is relatively meaningless. Based on this information, even though age might not be seen as a concern, sample size for some of age groups is a limitation of the current study.

Results regarding item analyses of the TV-GARS-2 showed that the test items provide ample evidence for the content-description validity. Although criterion-validity of the TV-GARS-2 could not be checked since there is no standardized assessment tool in Turkey, construct validity data also provided sufficient evidence for the Turkish version of the GARS-2. In the related literature, there are different criteria of acceptable item-discrimination coefficients regarding construct validity expressed by different professionals. For example, while Oosterhof (2001) stated that item-discrimination coefficients between .30 and .50 were acceptable, Özcelik (1989) pointed out that items having between .20 and .30 item-discrimination coefficients should be kept at the test. In addition, according to some professionals, items between .20 and .40 item-discrimination coefficients should be considered as acceptable items (Şencan, 2005). In our study, as can be seen at Table 1, almost all items are at acceptable levels except for items 2 and 24. Although item 2 and 24 have item-discrimination coefficients below acceptable levels, these items are being considered as important items and should be kept in the test by us since they are consistent with the structure

of GARS-2 as reflecting autism characteristics. In addition to item-discrimination coefficients, for example, age was found to have no meaningful relationship with TV-GARS-2. Using Hopkins's criteria, the correlation coefficients among subscales were found to be moderate to large in magnitude. Moreover, when examining the relationship of subscale standard scores with the Autism Index, significant correlations large in magnitude were found. This result supports that Autism Index can be used as the most valid predictor of the likelihood of the diagnosis of autism. Results of confirmatory-factor analysis also supported that construct of original GARS-2 was confirmed with Turkish sample. Regarding discrimination-validity of the TV-GARS-2, results revealed that TV-GARS-2 can be used in differentiating persons with autism from other special needs groups or persons with normal development and the identification of individuals with autism from individuals from other diagnostic groups.

Results related to internal reliability of the total scale and all the subscales of the TV-GARS-2 suggest that the items in the subscales are quite consistent and all of the subscales are sufficiently reliable. Test-retest data also suggested that TV-GARS-2 yields results that are stable over time. So, this information provides confidence in the subscales for professionals when making decisions or interpreting results from the TV-GARS-2.

The overall reliability and validity results of the TV-GARS-2 are very similar to those coming from other studies (e.g., Al Jabery, 2008; Gilliam, 2005) and similar results with the study carried with Greek population (Tafiadis, Loli, Tsanousa, & Tafiadi, 2008). These results support the knowledge base that autism is a universal disorder regardless of cultures. It also provides evidence that GARS-2 is a culturally robust scale. Overall, results of the study showed that TV-GARS-2 is a valid and reliable scale that can be used in difference settings.

#### References

- Al Jabery, M. A. (2008). *The examination of validity and reliability indicators of the Jordanian translated Arabic version of the Gilliam Autism Rating Scale (GARS-2)*. Unpublished doctoral dissertation, Wayne State University, Michigan.
- American Academy of Pediatrics. (2001). Committee on children with disabilities: the pediatrician's role in the diagnosis and management of autistic spectrum disorder in children. *Pediatrics*, 107(5), 1221-1226
- American Psychiatry Association (1994). *Diagnostic and statistical manual of mental disorders (4th ed.)*. Washington, DC: American Psychiatry Association.
- American Psychiatry Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed. Text Revision)*. Washington, DC: American Psychiatry Association.
- Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., & Chairman, T. (2006) Prevalence of disorders of the autism spectrum in a population cohort children in South Thames: The special needs and autism project (SNAP), *Lancet*, 368(9531), 210-215.
- Centers for Disease Control and Prevention (2012). New data on Autism spectrum Disorders. Retrieved from [http://www.cdc.gov/Features/Counting\\_Autism/](http://www.cdc.gov/Features/Counting_Autism/) on September 27, 2012.
- Çokluk, Ö. Şekercioğlu, G., & Büyüköztürk, Ş. (2010). Çok Değişkenli İstatistik SPSS ve LISREL Uygulamaları. Ankara: PEGEM.
- Diken, İ. H., Gilliam, J., Ardic, A., Diken, Ö., & Sweeney, D (2012). Results of pilot validity and reliability study of Turkish version of Gilliam Autism Rating Scale-2 (TV-GARS-2). *International Journal of Special Education*, 27(2), 207-215.
- Ellefsen, A., Kampmann, H., Billstedt, E., Gillberg, I.C., & Gillberg, C. (2007). Autism in the Faroe Islands: An epidemiological study, *Journal of Autism and Developmental Disorders*, 33(3), 703-707.

- Filipek, P.A., Accardo, P.J., Ashwal, S., Baranek, G.T., Cook, E.H. Jr., Dawson, G., (1999). The screening and diagnosis of autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, 29, 439-484.
- Filipek, P.A., Accardo, P.J., Ashwal, S., Baranek, G.T., Cook, E.H. Jr., & Dawson, G., (2000). Practice parameter: Screening and diagnosis of autism: report of the quality standards subcommittee of the American Academy of Neurology and the Child Neurology Society. *Neurology*, 55, 468-479.
- Gilbert, C., Cederlund, M., Lamberg, K., & Zeijlon, C. (2006). The autism epidemic: The registered prevalence of autism in a Swedish urban area. *Journal of Autism and Developmental Disorders*, 36(6), 429-435.
- Gilliam, J. E. (1995). *Gilliam autism rating scale (GARS)*. Austin, TX: Pro-Ed.
- Gilliam, J.E., (2005). *Gilliam autism rating scale: Second edition*. Austin, TX: PRO-ED.
- Hammill, D. D., Brown, L., & Bryant, B. R. (1992). *A consumer's guide to tests in print* (2nd ed.). Austin, TX: Pro-Ed.
- Hodge, D. (2008). *Children's sleep problems and maternal mental health in mothers of children with and without Autism* [Unpublished doctoral dissertation]. Claremont Graduate University, California.
- <http://www.cdc.gov/ncbddd/autism/index.html>
- İncekaş, S. (2009). Çocukluk Otizmini Değerlendirme Ölçeği Türkçe Formunun Geçerlik ve Güvenirlik Çalışması [Validity and reliability studies of Turkish form of Childhood Autism- Rating Scale-CAR]. *Unpublished doctoral dissertation*, 9 Eylül University Medical School, İzmir, Turkey.
- Irmak, T. Y., Sütçü, S. T., Aydın, A., & Sorias, O. (2007). Otizm Davranış Listesinin (ABC) Geçerlik Ve Güvenirliğinin İncelenmesi [Investigating reliability and validity of Autism Behavior Checklist (ABC)]. *Çocuk ve Ruh Sağlığı Dergisi*, 14(1), 13-23.
- Johnson, C.P., Myers, S.M., & Council on Children with Disabilities. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics*, 120(5), 1183-1215. Retrieved from [www.aap.org/pressroom/AutismID.pdf](http://www.aap.org/pressroom/AutismID.pdf) on November 16, 2009,
- Kline, P. (2000). *An easy guide to factor analysis*. New York: Routledge.
- Lecavalier, L. (2005). An evaluation of the Gilliam Autism Rating Scale. *Journal of Autism and Developmental Disorders*, 35(6), 795-805.
- Mazefsky, C., & Oswald, D. (2006). The discriminative ability and diagnostic utility of the ADOS-G, ADI-R, and GARS for children in a clinical setting. *Autism*, 10(6), 533-549.
- Oosterhof, A. (2001). *Classroom applications of educational measurement*. Englewood Cliffs, NJ: Macmillan Publishing Company.
- Özçelik, D. A. (1989). Test hazırlama kılavuzu [Guide for test preparation]. Ankara: ÖSYM Education Pub.
- Phillips, N. (2009). *A survey of the work and interactions of school psychologists, school neuropsychologists and clinical psychologists in schools today*. Unpublished doctoral dissertation, Claremont Graduate University, California.
- Schreck, K. A., & Mulick, J. A. (2000). Parental report of sleep problems in children with Autism. *Journal of Autism and Developmental Disorders*, 30(2), 127-135.
- Şencan, H. (2005). Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlilik [Reliability and validity in social and behavioral assessments]. Ankara: Seçkin.

- South, M., Williams, B. J., McMahon, W. M., Owley, T., Filipek, P. A., Shernoff, E., Corsello, C., Lainhart, J. E., Landa, R., & Ozonoff, S (2002). Utility of the Gilliam Autism Rating Scale in research and clinical populations. *Journal of Autism and Developmental Disorders*, 32(6), 593–599.
- Stone, W.L., & DiGeronimo, T.F. (2006). *Does my child have autism? : a parent's guide to early detection and intervention in autism spectrum disorders*. Jossey-Bass: San Francisco.
- Sucuoğlu, B., Akkök, F., & Gökler, B. (1996). Otistik Çocukların Değerlendirilmesinde Kullanılan Ölçeklere İlişkin Bir Çalışma [A study regarding scales being used in the process of assessment of children with autism]. *3 P Dergisi*, 4(2), 116-121.
- Tafiadis, D., Loli, G., Tsanousa, E., & Tafiadi, M. (2008). *The Gilliam Autism Rating Scale (GARS - 2), a pilot study for the Greek autistic population*. Poster presented at International Society on Brain and Behaviour: 3rd International Congress on Brain and Behaviour, Thessaloniki, Greece. 28 November – 2 December 2007.
- Tetik Kabil, A. (2005). Türk örnekleme ile yapılan Erken Çocukluk Dönemi Otizm Tarama Ölçeği geçerlik çalışması [Validity study of the Checklist for Autism in Toddlers- CHAT with Turkish sample]. Unpublished master's thesis, Boğaziçi University, İstanbul, Turkey.
- Trembath, D. Balandin, S., & Rossi, C. (2005). Cross-cultural practice and autism. *Journal of Intellectual & Developmental Disability*, 30, 240-242.
- Ünal, F. Pehlivan Türk, B. Özusta, Ş. Şimşek, Ç. Çalışgan, L. Samancı, Ş. & Münir, K. (2006). Otizm Tarama Çalışması Sonuçları ve Değiştirilmiş-Oyun Çocuğunda Otizm Tarama Listesinin Psikometrik Özellikleri [Results of autism screening study and psychometric characteristics of the Modified Checklist for Autism in Toddlers-M-CHAT]. Paper presented at the 16<sup>th</sup> Child and Youth Mental Health and Diseases Congress, Antalya, Turkey.
- Volkmar, F.R., Cook, E.H., Pomeroy, J., Realmuto, G., & Tanguay, P. (1999). Practice parameters for the assessment and treatment of children, adolescents, and adults with autism and other pervasive developmental disorders. American Academy of Child and Adolescent Psychiatry Working Group on Quality Issues. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38(12), 325–545.
- Volkmar, F.R., Klin, A., Siegal, B., Szatmari, P., Lord, C., & Campbell, M., (1994). Field trial for autism in DSM-IV. *American Journal of Psychiatry*, 151, 1361-1367.
- World Health Organization (1993). *The ICD-10 classification of mental and behavioral disorders: Clinical descriptions and diagnostic guidelines*. Geneva, Switzerland: World Health Organization.
- Yıkgeç, A. (2005). Değiştirilmiş Erken Çocukluk Dönemi Otizm Tarama Ölçeğinin Türk Örnekleme Üzerinde Geçerlik Çalışması (*Validity study of the Modified Checklist for Autism in Toddlers-M-CHAT with Turkish sample*). Unpublished master's thesis, Boğaziçi University, İstanbul, Turkey.