Veysel Aksoy [1]

İbrahim H. Diken [2]

Identification of Psychometric Characteristics of Sample of Vocal Behaviour and Interaction Assessment Record Forms*

Abstract

The purpose of this research is to study the Sample of Vocal Behaviour Record Form (SVBRF), included in the third edition of Autism Screening Instrument for Educational Planning-3 (ASIEP-3), and the psychometric properties of the trucked forms for the informal assessment tools of Autism Screening Instrument for Educational Planning-3 (ASIEP-3). The rationale for the research is the lack of standardised informal assessment tools to use for the educational assessment of children with Autism Spectrum Disorder (ASD) in Turkey. The study group consists of students, who receive education at three private institutions for special education and rehabilitation. A total of 169 students were included in the study group of the Turkish Version of the SVBRF, 54 females and 115 males. And in the study group of the Turkish Version of the IARF, a total of 115 students were included, 37 females and 78 males. The data, which were obtained from the study requiring data collection based on performance feedback, were analysed in order to assess the validity and reliability of the tools. The findings from the analysis revealed that both tools were valid and reliable to be used for obtaining the samples of vocal behaviour of children with ASD, and evaluating their level of social interaction.

Keywords:
Autism Spectrum Disorder, Informal Assessment, Sample of Vocal Behaviour, Social Interaction

Introduction

Autism Spectrum Disorder (ASD) is considered as a neurodevelopmental disorder and one of the special education category, also typically appears during the first three years of life, having become increasingly prevalent in the recent years. ASD affects language, cognitive and social development of children, and compared to other types of disorders, the prevalence of ASD has unexpectedly increased during the last 20 years (Allen, Robins and Decker, 2008). As in many other parts of the world, the exact number of children with autism is not known in Turkey. Although the prevalence rates of autism range from 1 out of every 100 to 160 children according to official estimations, it would not be possible to provide exact figures (Glosser, 2011). According to the data on the prevalence of autism released in 2012 in the United States, being one of the leading countries where the records on this field have been kept for longer, 1 in 68 children has been identified as having an autism spectrum disorder (CDC, 2016). Due to the increase in the prevalence of ASD, there has been an increasing awareness of this disorder generally in the society particularly among parents, healthcare and educational professionals and researchers (Dover ve Le Couteur, 2007).

[1] Ph.D., Asst. Prof. Dr., Anadolu University, Research Institute for Individuals with Disabilities, Eskişehir, TURKEY.
e-mail: vaksoy@anadolu.edu.tr
[2] Ph.D., Prof. Dr., Anadolu University, Research Institute for Individuals with Disabilities, Eskişehir, TURKEY.
e-mail: ihdiken@anadolu.edu.tr
*This study was conducted based on the first writer's doctorate study and the supervision of the second writer.

Although symptoms of ASD vary considerably in each child, three core symptoms are viewed as common problematic areas: communication, social interaction, and repetitive behaviours and restricted interest. The fact that severity and level of these problems vary in each individual has had influence on the fact that ASD has become a disorder classified under an umbrella category. According to the Diagnostic and Statistical Manual of Mental Disorders-IV-TR (DSM-IV-TR, 2001), released by the American Psychiatric Association (APA), there are five sub-categories under this umbrella: Asperger's Syndrome, Autism, Rett's Syndrome, Childhood Disintegrative Disorder, Pervasive Developmental Disorder not Otherwise Specified which includes atypical autism. As referred in DSM-V, published by the APA in May 2013, this umbrella category was replaced with a single term: Autism Spectrum Disorder (ASD).

Assessing children with ASD was defined as a process of gathering data/information from various sources using various methods and tools in order to determine development of children, precautions that should be taken and necessary interventions (Sucuoğlu, 2012). A similar process is followed for the assessment of children with ASD as in the other types of special needs.

And finally, the assessment of children with ASD in terms of special needs services can be viewed as a data gathering process to discover academic, behavioural or physical characteristics of pupils, and make legal and educational decisions applicable to these characteristics (Gürsel, 2012). Sucuoğlu (2012) states that the assessment phases of individuals with ASD can be studied in three groups as screening, diagnosis and gathering developmental data. The phase of gathering developmental data requires educational evaluation. Tincani (2010) states that diagnosis should be supported with a developmental assessment during the phase of diagnosis. Tincani also points out that the scores obtained from the diagnosis tools during the phase of diagnosing children with ASD will not be sufficient enough to make a definitive diagnosis, and that children should also be assessed developmentally, particularly in such areas as language-communication and social interaction in order to make a definitive diagnosis. A multi-disciplinary team should carry out this assessment in various environments using multiple data sources.

Avcıoğlu (2011) indicates that following a diagnosis, a comprehensive educational assessment should be carried out in order to make relevant legal and educational decisions for the students, whose eligibility has been determined for special needs education, plan educational intervention programs and then measure the efficiency of interventions. Peterson and Meier (1988) stated that a child's level of functioning in various developmental areas is determined, and level of performance is presented for an Individualized Education Program (IEP) developed for each child. Thus, strengths and weaknesses of a child, and what types of interventions are needed in which areas are identified. Within the scope of an educational assessment, mainly communicational and language skills, social and adaptive behaviours, cognitive and academic skills of children with ASD are assessed, and formal and informal assessment tools can be used during this phase (Sucuoğlu, 2012).

As is known, formal assessments are norm referenced and standardised assessments. The tools used during the phases of screening and diagnosing ASD both globally and in Turkey have these characteristics. The assessment tools, which are used to obtain clinical findings and conduct assessment, are expected to have these characteristics. However, it is not always possible to make a definitive ASD diagnosis. Therefore, the fact that assessments are supported with evaluations covering the areas of particularly language-communication and social interaction will contribute to the diagnosis and planning an intervention program to be developed for a child.

The two tools, whose psychometric properties were studied and adapted into within this study, are two informal tools included in the third edition of the Autism Screening Instrument for Educational Planning-3 (ASIEP-3) which was developed by Krug, Arick and Almond (2008). The ASIEP-3 tool set is an Autism Behaviour Checklist (ABC) that indicates children with a probability of ASD, consisting of four informal assessment record forms standardized through the outputs obtained from Intellectually Disabled (ID) individuals. ABC, contained in the tool, is a norm referenced screening instrument. It was stated that the ABC could be used for

screening and diagnosing specifically children suspected of having ASD. The other four tools, included in this set and studied for their psychometric properties as part of this study, were standardised informal assessment tools. Although the tools were already standardized, their norm references were not created in an extensive sampling.

Instead, there are sampling references, which enable a comparison with a diagnosed sampling group. Therefore, it is indicated that these tools are described as informal yet standardised tools, since their validity and reliability analyses were done, applied and scored, and interpreting their results required standard procedures (Krug, et al., 2008).

Krug et al. (2008) indicate that ASIEP-3 can be used for diagnosing children with ASD, placement, planning their educational programs and monitoring their development. Considering the existing diagnosing and screening tools in Turkey, it appears that we experience a lack of tools that enable an objective comparison of language- communication and social interaction performances of individuals with ASD to a sampling group. During post-diagnosis phases, the programs for students are planned through criterion-referenced assessments created by teachers, and student development is monitored along this program. These types of standardised informal tools allow for an objective measurement and assessment to be used during the phases of program development, program evaluation and monitoring for individuals with ASD. Thus, Turkish adaptation of the tools of Sample of Vocal Behaviour Record Form (SVBRF) and Interaction Assessment Record Form (IARF) included in ASIEP-3, and studying their psychometric properties were deemed important. In this regard, this study seeks answers to the following research questions:

- Is the Turkish Version of Sample of Vocal Behaviour Record Form (SVBRF-TV) a reliable and valid tool for the assessing vocal behaviours of children with ASD?
- Is the Turkish Version of Interaction Assessment Record Form (IARF-TV) a reliable and valid tool for the assessing social interaction level of children with ASD?

Methodology

Research Design

This research was conducted through a general screening model. General screening models can be applied on an entire universe or a sample taken from a universe in order to make a general judgement about a universe composed of many elements (Karasar, 2008). In accordance with the model, this research did both single and relational screening. Through the tools, whose psychometric properties were aimed to identify, both the types and quantities of each variable, and the relationships among the certain dimensions of the variables were studied.

In this study, children with ASD and ID were compared in terms of vocal behaviours and social interaction, and a causal-comparative research design was used. In causal-comparative researches, two or more groups of subjects (ASD and ID in this research) were matched in terms of certain characteristics, and compared in terms of one or more variables (vocal behaviour and social interaction in this research) (Gay, Mills and Airasian, 2006). Therefore, the model of this research was defined as a correlational and causal-comparative research model.

Participants

The participants of this research were the students diagnosed with ASD and ID, having special educational needs, and the teachers from three private institutions for special education and rehabilitation. Although the tools, whose psychometric properties were aimed to identify, will be used for assessing the children, the reason for obtaining data from the children with ID as well as ASD can be explained in the following way: First of all, data were obtained from both groups in the original study; therefore, as it was aimed to adhere to the original study during the study of identifying psychometric properties, a similar approached was performed.

Secondly, on the analyses of distinctiveness conducted in order to test the validity of the tool, obtaining data from the students with moderate and severe ID, the closest disability group to the ASD group for their characteristics, will contribute more factual proof to distinctiveness. Thirdly, special education professionals use similar methods and procedures for the educational evaluation of

children with both ASD and ID. When choosing the study group of the research, a purposive sampling technique, which is one of non-contingent sampling methods, was used. The research design requires accessing a certain number of children with ASD and ID between the ages of 2 and 20 in average. Moreover, since the application of research requires the teachers working with these children to be examiners, the criterion of teachers' volunteering was added to the criteria of diagnosis, age and number of students. The method of research application is practices based on interaction, having the requirements of specially designated environment, materials and time, requiring measuring and recording student performance for a length of time within semi-structured environments. And this is a factor, which makes the implementation difficult in state schools.

Therefore, along with the characteristics of students and volunteering of teachers, the choice of sampling is included variables that can be defined as accessibility and applicability. Due to all these factors, private institutions for special education and rehabilitation were chosen as the implementation areas, while the sampling of the research was selected. The students attending these institutions and meeting the criteria were included in the sampling. Those with the diagnosis of high functioning autism and atypical autism among the students with the diagnosis of ASD were not included in the sampling. And among the non-ASD group, those with the diagnosis of moderate and severe ID were included in the sampling.

A total of 33 teachers, 26 (79%) females and 7 (21%) males were included in the research during the phase of data gathering for the tools, whose psychometric properties were studied. The teachers are in the age range of 22 to 57 with an average age of 33.8, standard deviation of 10.72 and a range of 35. The occupational experience of teachers range between 1 to 26 years with an average of 7.03 years, standard deviation of 5.77 and range of 25. The data about the students, whose data were collected within this study, are given on Table 1.

Table 1.

Data About the Children with Special Needs Participated in the Research

	Sex	**Age***						
		n	%	min.	max.	X	ss	Range
SVBRF-TV	ASD-Female	26	28	2;2	22;7	9;2	4;7	20;5
	ASD-Male	68	72	2;1	18	9;1	4;11	15;11
	Total	94	100	2;1	22;7	9;1	4;5	20;6
	Non-ASD-Female	28	37	2;2	19;10	10;11	4;2	17;8
	Non-ASD-Male	47	63	2;6	20	10;2	5	17;6
	Total	75	100	2;2	20	10;6	4;8	17;10
	General-Female	54	32	2;2	22;7	10;1	4;7	20;5
	General-Male	115	68	2;1	20	9;6	4;7	17;11
	General-Total	169	100	2;1	22;7	9;8	4;7	20;6
IARF-TV	ASD-Female	22	34	2;3	22;7	8;10	5;1	20;6
	ASD-Male	42	66	2;1	17;7	10;1	4;4	15;6
	Total	64	100	2;1	22;7	9;8	4;7	20;6
	Non-ASD-Female	15	29	2;2	14;6	9;10	3	12;4
	Non-ASD-Male	36	71	2;6	17;10	9;9	4;6	15;4
	Total	51	100	2;2	17;10	9;9	4;1	15;8
	General-Female	37	32	2;2	22;7	9;3	4;4	20;5
	General-Male	78	68	2;1	17;10	9;11	4;5	15;9
	General-Total	115	100	2;1	22;7	9;8	4;4	20;6

*Ages were given out in years and months, and the first figure represents year, and the second figure represents month in the punctuation mark of ";" (e.g., 12;7 - 12 is age, and 7 is month of the birth date).

Instruments and Procedures

Since this research is the phase of identifying the psychometric properties of assessment tools, the validity and reliability findings related to the scores of assessment tools obtained from the research were reported in detail in the Findings section of the research. Thus, in this section, the validity and reliability findings reported in the development of the original form of the tool, and in the following section, the process was reported.

Sample of Vocal Behaviour Record Form-Turkish Version-(SVBRF-TV)

The purpose of the tool of Sample of Vocal Behaviour Record Form-Turkish Version (SVBRF-TV) is to record the language sample, which reflects the vocal expression of chid in the best way, even if the utterance of child is limited to only crying, coughing, laughing and giggling. SVBRF-TV is used in order to evaluate the level of expressive language during the preverbal stage, and when talking skills begins. This component measures the four characteristics of natural (spontaneous) talking skills of children with ASD. These skills are repetitive (stereotypical), non- communicative (related to social interaction), unintelligible (delay and deviation in expressive language) and babbling (meaningless vocalisation) characteristics. This tool is also used for monitoring the level of learning verbal language for children with ASD (Krug and et al., 2008).

The materials required for the use of this tool are data form, paper, pen, timekeeper and optionally a voice recorder. The areas for applying this tool are playground, leisure area or an unstructured environment such as an activity table. The examiner of the tool can both bring out utterances and record them at the same time, or another examiner can record while the examiner brings out the utterances. It is recommended that a voice recorder, although it is not compulsory, should be used during the implementations when there is a single examiner. The examiner interacts with child without giving direct instructions in order to bring out utterances from child while passing stimuli, such as a book, toy or physical stimuli. Since the purpose of this activity is to record the language skill of child, primary reinforces used in structured instructions and tutorial sessions are not provided to child. This procedure is ended when child responses with 50 utterances. If the session lasts more than 30 minutes and child still has not responded with 50 utterances, the session is ended and data gathering is continued on the second session conducted on the same or following day. This practice ends when child reaches 50 utterances on the second session. If child cannot reach 50 utterances in the second session either, a third session is conducted. So, the maximum length of this test is 30+30 = 60 minutes.

There are various scoring categories with each of which helping to display the developmental variances in language acquisition. These categories are *variety, function, articulation* and *length.* After utterances are recorded, each utterance is scored for each category. *Variety* shows if a verbal expression is the repetition of the previous expression or used for the first time; and *function* displays verbal expressions, which are and are not communicational.

Articulation means the intelligibility of a vocal expression. *Length* means the number of words in an utterance, and the column of length also includes a space for recording sounds articulated except from words (for instance, pre-babbling or babbling). *Vocalisation Complexity* and *Parts of Speech* columns were added in order to gather additional diagnostic data to be used for planning the educational program of children. Scores for autistic speech characteristics and normal language development are obtained by scoring these sections.

Apart from this section, there is an *Optional Analysis* section that meets the needs of speech and language experts, and enables analysing the obtained language sample in detail. *Optional Analysis* enables the child's language sample to be analysed according to the *Vocalisation Complexity and Parts of Speech* characteristics. Child's words are analysed under the titles of *Isolated sound, C-V Blend, Simple 2-Syllable Combination, V-C Blend, C-V-C Blend* and *Complex Multi-Syllable Combination*, and their percentage distribution can be obtained on the section of *Vocalisation Complexity*. And on the *Parts of Speech* section, only intelligible words are

scored. Conformity to the rules of grammar and complexity of words are scored again on this section. The section of *Optional Analysis* is a detailed section of analysis enabling therapists to work on child's language and speech skills to record individual planning and progress. A statistical analysis has not been reported regarding this section.

During the original standardization of SVBRF, 81 examiners conducted the application study through the data obtained from 157 participants. Participants included 61 children with ASD and 96 non-ASD children (mental deficiency). The age range in the sampling is 2 years and 4 months to 20 years old.

The reported reliability-validity findings of SVBRF are as following: Three kinds of reliability data findings were reported as alpha reliability coefficient, estimated with two split-half technique, test-retest technique and inter-observer recording reliability of the participant on the reliability stage. Spearman-Brown prediction formula was used for the estimated reliability coefficient from the entire test for the two semi-test reliability coefficients.

The results of this test are r (59) = .97, p < .01. For test-retest reliability, Pearson correlation coefficient was calculated with the data obtained at three-day intervals from 20 participants. The correlation coefficients of the test-retest reliability for each sub category are as following: Repetitive is .95; non-communicative is .92; Babbling is .81, First Use .85 and Communicative-Intelligible is .94. The average reliability for each autistic speech characteristics (repetitive, non- communicative, unintelligible and babbling) on the interrater recording reliability was calculated as 90%.

On the level of Validity, the findings of the Content Description Validity, Criterion Prediction Validity and Construct Identification Validity were given. The researchers reported their qualitative justifying based on expert opinion in accordance with a literature review on the level of Content Description Validity. And on the level of Criterion Prediction Validity, the correlations between the SVBRF and the Sequenced Inventory of Communication Development/SCID language scores, developed by Hedrick and Tobin (1975), were analysed.

The Pearson correlation (r = .81, p < .01, n = 83) was found among the total scores. Secondly, the total for the score of Expressive Language Age of SCID and SVBRF's raw scores of First Use and Communicative Use were compared, and (r = .82, p < .01, n = 27) was calculated as the result. And thirdly, the correlations with the sub-dimensions of Receptive Language, Expressive Language and Communication Domain of the Vineland Adaptive Behaviour Scales: Interview Edition were referred (n=133) through the SVBRF scores. On the results of this analysis, it was reported that there was a correlation in the range of -.23 and .75 among the sub-dimensions of Receptive Language, Expressive Language and Communication Domain of the Vineland Adaptive Behaviour Scales: Interview Edition. On the dimension of Construct-Identification Validity, the correlations among the sub-domains of autistic speech scores of the SVBRF were referred (n=139). And it was reported that the coefficient of correlation obtained from this varied between .52 and .81 (Krug et al., 2008).

Interaction Assessment Record Form- Turkish Version-(IARF-TV)

Interaction Assessment Record Form is used in order to generate social interaction behaviours of individuals within semi-structured play environments in which systematically developed stimuli are presented. This record form is used for both assessing spontaneously occurring social interaction behaviours and responses demanded by the examiner. A behaviour sample, which can represent the behaviour exhibited by the child during a certain time length, is obtained at the end of this practice.

While one person interacts with the child in a play environment, another person codes the interaction in order to evaluate a child using IARF-TV. A time sampling record is used in coding interaction. A voice, recorded to a CD or a voice recorder, speaks out when monitoring begins and ends. Each monitoring session is coded in one of the cells on the second page of IARF-TV. There is a matrix format that shows what the adult in the environment is doing before the child's

response, and the description of the behaviour corresponding the child's response in the form. While the adult behaves in accordance with the order of the situations given in the form, the child's behaviour is recorded in the columns as monitoring number by the examiner. What occurs in the end is not the frequency count of various behaviours, but a sample related to the child's general behaviour. At the end of this practice, the child's scores are recorded, and the percentage value of each defined behavioural category in the interaction is calculated. Also, the autistic interaction score is obtained by substracting the child's interaction score from 48 and adding the result unresponsiveness score to it. The level of the child's autistic interaction can be checked against the norm of sampling group by comparing this score to the given percentile table.

This procedure is carried out in a play environment where the child can interact with the adult. The behaviours of the child can be monitored in an area of approximately 3x3 meter, where there are various toys suitable for the child's level of development. While an adult interacts with a child, an observer monitors and records the child's behaviours according to time sampling. In the time, sampling record, the observer monitors the child for 10 seconds, and on the 10th second, records the number he or she hears from the voice recorder in the matrix box corresponding the child's behaviour at that moment in 5 seconds. Apart from the arranging of the environment, a voice recorder, a form, paper and pen are needed for this practice. This practice consists of three stages each of which lasts 4 minutes, and the total length of monitoring is 12 minutes. During this procedure, the adult in the environment has three behaviour types defined for each stage. These are *Active Modelling, Passive-No Intervention and Direct Clues.*

During the stage of *Active Modelling,* the behaviours that the adult needs to display in the play environment are defined as following: The adult sets an example that the child will also can do. The adult plays with the toys invitingly, and explains what he or she does verbally to the child. He or she does not directly invite the child to the play and say sentences such as "*Look! Wouldn't you want to pile the cubes?*" "*Isn't this fun?*" etc., as these expressions requires the child to give a response. While the adult continuously tells what he or she does or talks about the toys, he or she says, "*I have fun while I pile the cubes*," and makes the sound of "*vroom vroom*" while driving a toy car or voices two puppets. If the child doesn't get involved in the play, no clues or directions should be given. Any toys shouldn't be passed to the child, because these kinds of behaviours are direct clues, and require giving response. If the child starts an interaction at any moment (for instance, holding the adult's hand, asking the adult to do something, and playing with or close to the adult), the adult will respond the child's request in a normal way, but not continue the interaction. He or she simply does what the child asks, and goes back to the modelling activity. During the stage of *Passive-No Intervention,* the behaviours that the adult needs to display in the play environment are defined as following: The adult ends his or her behaviour of being a model in the previous stage, and sits on a chair or the floor. He or she does not become a model, describe or talk, just sits passively and waits.

If the child starts an interaction at any moment (for instance, holding the adult's hand, asking the adult to do something, and playing with or close to the adult), the adult will respond the child's request in a normal way, but not continue the interaction. What is assessed here is how the child interacts with the adult, and starting or continuing an interaction is up to the child. The adult does what is required by the child, and goes back to where he or she is before.

During the stage of *Direct Clues,* the behaviours that the adult needs to display in the play environment are defined as following: The examiner gives a direction approximately in every 10 seconds to add up to total 6 directions in every minute. He or she can determine these directions beforehand. In order to figure on the ten second intervals, he or she can make use of a big clock hung on the wall. Directions should be given from a distance, not too close or far away from the child, and with an encouraging tone of

voice. The examiner can use symbols or gestures, however, should not use full physical clues, which will bring out correct responds. Here, behaving appropriately is not the target; therefore, the child should not be supported for him or her to carry out the instructions. At this stage, child's possible responds were classified, and four types of child responds were defined. These are Interaction, Constructive Independent Play, Unresponsiveness and Negative Aggressive.

If the child's behaviour is explicitly agreeing with the fact that there is an adult in the environment, the column of *Interaction* is coded. If the child reacts to the adult's question, direction or clue (properly or improperly), starts an interaction and/or continues the interaction (exchange with the adult), follows the instructions, mimics or copies a model given in the place, touches the adult physically or reaches out, communicates through gestures, signs or talking, his or her behaviours are coded in the interaction column. Eye contact is not enough on its own for the behaviour to be considered an interaction, as overt behaviour is needed for interaction.

The *Constructive Independent Play* column is coded when the child plays on his/her own and is not in interact with the adult. For the response to be coded as an independent play, the child needs to be in a type of independent play. The behaviours below are described as examples for constructive independent behaviours.

- child sitting facing back to the adult and plays with a toy
- child continuing to play with the toy as previously although the adult gives an instruction
- child quietly diving into the toy box
- child pulling a rope, pushing the toy trucks, placing cubes on top of each other or rolling a ball.

The behaviours scored in the *No Response* column involve isolated behaviours. If the child:

- has repetitive behaviours with no apparent reasons unless he or she needs an emotional outlet (repetitive games such as twisting rope, crossing fingers, spinning and repeatedly spinning the wheels on toy cars) - (*Self-stimulation)*
- bites, scratches himself/herself or bangs head on the wall - (*Self-injury)*
- screams, shouts or throws himself on the floor with no apparent reason - (*Uncontrollable tantrum)*
- cannot demonstrate observable behaviours, 'no response' column is marked - (*Lack of observable behaviour)*

The *Aggressive Negative* column indicates verbal expressions of non-verbal actions that show the child's anger, rejection and lack of courage towards the adult in the environment. Aggressive Negative can be vocal as well as non-vocal:

- Non-vocal behaviours include (a) child's anger directed after a direct clue, (b) that the child hits the adult in the environment, (c) pushes the adult or (d) throws an object at the adult.
- Vocal behaviours should follow that the child (a) demonstrates directed anger along with shouting, (b) violent rejection (such as, shouting "No!" - for verbal rejection to be scored as negative, it either needs to follow an instruction or an); (c) verbal abuse (such as swearing, calling bad nicknames) or (d) directed yelling or whining.

During the original standardization of IARF-TV, 60 examiners conducted the application study through the data obtained from 115 participants. Participants included 52 children with ASD and 63 non-ASD children. The age range in the sampling was not reported. The interjudge reliability coefficient on the level of reliability calculated with the Kuder-Richardson item reliability, and the result was r = 0.86 (n=87).

Along with the content-description validity based on literature and expert opinion on the level of Validity, Construct-Identification Validity and subgroup variations analyses were conducted. On the level of Construct-Identification Validity, it was reported that those, who got high scores on the level of No Response of IARF-TV, got high scores also on the Autism Behaviour Checklist/ABC (r = 0.61, $p < .01$). On the analyses of distinctiveness made with 23 children with ASD and 23 non-ASD children, it was reported that IARF-TV distinguishes children with

ASD from non-ASD children. The reported *F* values vary between 6.7 and 37.7, and between p < .01 and .05 (Krug and et al., 2008).

Early Language Development in Turkish (TEDIL)

TEDIL is the Turkish adaptation of TELD-3, which is an assessment tool for the receptive and expressive language skills of the children between the ages of 2 and 8 years, and was developed by Hresko, Reid and Hammill (1999). TEDIL test set includes the following compounds: TEDIL picture book, Form A Examiner's Record Form, Form B Examiner's Record Form, TEDIL equipment - A doll, a pair of baby shoes, a toy car, a small ball, a spoon, 5 coins (kurus/lira), 5 cubes, examiner's user guide.

TEDIL consists of two parallel assessment tools as *A* and *B*. The picture manuals and application record forms of each assessment tool are separate. Each set includes two sub-tests: Receptive and expressive. There are 76 articles which assess the semantics and grammar fields of language in each parallel form. Some articles have sub-articles. Each receptive language sub-test includes 37 articles. Form A is composed of 24 articles that assess semantics and 13 articles that assess grammar. There are 39 articles in each of the two expressive language sub-tests. Form A is composed of 17 articles which assess semantics and grammar. TEDIL A Form was used in order to perform criterion-referenced validity analyses. The reported Cronbach Alpha coefficients of the test vary between .86 and .98 for the receptive language A and B forms. Expressive language A form range between .87 and .98, and B form between .86 and .98. Test retest-reliability is .96 for the Receptive Language A Form, .93 for B Form, .89 for Expressive Language A Form, and .83 for B Form (Güven, 2009; Güven and Topbaş, 2009).

Gilliam Autism Rating Scale-Turkish Version-2/GARS-2-TV

On this research, data were collected through 2/GARS-2-TV in order to use on the criterion-referenced validity analyses of the four developmental assessment tools of the ASIEP-3 that was adapted to Turkish. 2/GARS-2-TV is an assessment tool developed to collect data for screening/diagnosing individuals with ASD, assessing severe behavioural problems, identifying/assessing behavioural improvement, determining goals for individualised education program and scientific researches. This tool is individually applied to individuals suspected of having ASD between the ages of 2 and 23. There are 3 subscales which include 42 articles in total. Subscales are *Stereotyped Behaviours* (14 articles), *Communication* (14 articles) and *Social Interaction* (14 articles). 2/GARS-2-TV is scored through a Likert type scale. Never observed (0 point), rarely observed (1 point), occasionally observed (2 points) and frequently observed (3 points). The collected points can be turned into standard points and percentile rank. Interpretations of the points can help guess the likelihood of having ASD. A researcher is a person who had the training to use this tool and is capable of applying it.

The adaptation of the Turkish version of this tool has been done with the data obtained from 1,191 children with ASD all around Turkey. Along with a linguistic equivalence study, findings on the validity and reliability were reported. Cronbach's Alpha Coefficients for each subscale on the level of internal consistency are .98 for the subscale of Stereotyped Behaviours, .99 for the subscale of Communication and .99 for the subscale of Social Interaction. Cronbach's Alpha Coefficient for a total of scales is .99. Test-retest reliability is .98 for the subscale of Stereotyped Behaviours, .99 for the subscale of Communication and .99 for Social Interaction. The total scale is .99. On the analyses conducted by testing the hypotheses in the original scale for the validity, 6 basic hypotheses, used during the validity study of the original scale, were tested. It was reported that all the hypotheses were proven to be true on the validity analyses (Diken, Ardıç, Diken ve Gilliam, 2012).

The Process

Before we started the research, we had obtained all the necessary permissions to use the copyrighted tools, whose psychometric characteristics were aimed to identify, from the US organisation holding the copyrights.

The manual for the use of tools, which will be included in the study, was translated into Turkish to get detailed information about the use of each tool. The Turkish version of the original forms of the tools was done through translation and retranslation. After the tools are translated into Turkish, materials needed for the implementation of each tool were obtained. For the implementations, the staff in these organisations, and teachers volunteered to participate in this research were trained on the introduction, implementation of the tools, and data collection. Teachers performed the implementations and collected data. The average duration of training for each tool was an hour. First of all, introduction, use and scoring of these tools were presented on these trainings. Then, the examiner demonstrated how the implementation would be performed as a model. On the last part of the training, all the teachers shared their views on the implementation of the tool, and asked questions interactively. Training sessions continued until the teachers stated that they had the sufficient amount of information regarding the use of the tools.

6 training meetings for the examiners were held in total with 33 teachers, who collected data during the research, including 3 meetings for each tool. The performance data was collected through 284 sessions in total including 169 (157 in the original study) for SVBRF-TV, and 115 (115 in the original study) for IARF-TV. For Criterion Prediction Validity, 2/GARS-2-TV data were collected from 61 students, and TEDIL data from 31 students. Also, 33 test re-test data were collected from SVBRF-TV. When all the tests were performed together, 409 measurements were conducted.

Data Analysis

Before data analysis, each data form was reviewed to check whether there were any forms filled out incorrectly. During the process of data collection, the researcher was continuously in contact with the teachers who performed the implementation; therefore, no omissions or errors were given in the study. As children's date of birth/name sections were not filled on some data forms, the researcher filled the missing sections by checking with the information on the records, and staff of the institutions where the children attend to. Although data were collected, total scores were not entered on the front page of some data forms, and the forms were prepared for filling in the data set of the statistics program on the electronic environment. On the analyses conducted on the data files created on the statistics program, the assumptions, which are considered as prerequisites of the related statistics, were tested before the parametric analyses were conducted. First, descriptive analyses of the participants were conducted on the related analyses of the two tools, whose reliability and validity reviewed within the scope of the research. Pearson correlation coefficient was calculated for the test-retest reliability that was done in two weeks interval. The rate of inter-observer reliability was calculated as a percentage for inter-rater reliability.

Correlation coefficients of inter sublevels on the reliability analyses were calculated with the Pearson correlation coefficient. The levels of distinguishing ASD from non-ASD children for the rating scale were created using t-test. Moreover, criterion-referenced validity of the scales was analysed through calculating Pearson correlation coefficient. SSPS software package was used for statistics calculations.

Results

Reliability-Validity Findings Related to Scores Obtained from SVBRF-TV

On the original study, the methods of two semi-tests, a test-retest and interrater recording reliability were used in order to examine the reliability of SVBRF-TV. And on the level of validity, the proof of criterion-referenced and construct-identification validity was presented.

In this study, the reliability and validity of this tool, which was translated into Turkish by using the same method, were examined, and the Cronbach Alpha coefficient for the four items forming the autistic speech score and item-total score correlation values were calculated.

Internal consistency coefficient

The test was divided into two as Test A and Test B on the basis of the Autistic Speech Score that the participants obtained from SVBRF-TV for the reliability of the two semi-tests, and the reliability of the test was calculated through the relationship between the two semi-tests. The results were r (169) = .86, p < .01. This figure is the reliability coefficient calculated for one half of the test. The result for the reliability coefficient of the entire test, which was calculated using the Spearman-Brown prediction formula, was r (169) = .92, p < .01. The findings reported on the original study were r (59) = .95, p < .01 for one half of the test, and r (59) = .97, p < .01 for the entire test.

Coefficient of Stability

The coefficient of stability for the test was calculated on the basis of the data obtained from 33 participants at two weeks intervals through the use of the Pearson correlation coefficient. The correlation coefficients of the test-retest obtained from for each category, and autistic speech category are .86 for *Repetitive*, .94 for *Noncommunicative*, .80 for *Unintelligible*, .82 for *Babbling* and .96 for *Autistic Speech Total Score*. On the category of the autistic speech total score, the results are .82 for *First Use*, .93 for *Communicative*, .90 for *Intelligible* and .96 for *total Speech.* The test-retest reliability on the original study was examined through the data obtained from 20 participants at three days intervals. However, the findings for *babbling* and *total score* from the autistic speech category, and for *intelligible* and *total score* from the total speech category were not reported. The reported findings are as follows: .95 for *Repetitive*, .92 for *Nonintelligible*, .81 for *Babbling*, .85 for *First Use*, and .94 for *Communicative*.

Interrater Reliability

For the interrater reliability input, data were collected together with the examiners on 16 implementations, and the overlap percentages of the two records on the levels of diversity, function and articulation were reviewed. The interrater reliability was calculated with the formula of [*Consensus / (Dissensus + Consensus) x 100*]. Interrater reliability was calculated as 88% on the level of diversity and function, 90% on the level of Articulation and 89% on average. The percentage values of the sublevels on the original study were not reported. The reported average of the reliability was 90%.

Item-total score correlation

Although it was not calculated on the original study, Cronbach's alpha coefficient was calculated as α = .92 for the internal consistency of SVBRF-TV on the basis of the data obtained for this study. The item-total correlation values of the four items making the autistic speech score were calculated as .78 for repetitive, .87 for noncommunicative, .85 for unintelligible and .79 for babbling.

Criterion-related validity

The correlation coefficient between the total speech scores of SBVRF-TV and TEDIL receptive language score is .71 for the analysis conducted on the basis of data obtained from 31 children and implemented by the use of both TEDIL and SVBRF-TV. The correlation coefficient between the total speech scores of SVBRTF-TV and TEDIL expressive language score is .79. The results were noteworthy on the level of p < .01. On the analysis, conducted through the data collected from 61 children with the use of both 2/GARS-2-TV and SVBRF-TV, the correlation coefficient between the autistic speech scores of SVBRF-TV and communication subscale score of 2/GARS-2-TV was calculated as .70; the correlation coefficient between the autistic speech scores of SVBRF-TV and social interaction subscale score was calculated as .49, and the correlation coefficient between the autistic speech scores of SVBRF-TV and stereotype behaviours subscale score was calculated as .50. The obtained figures are noteworthy on the level of p < .01. Criterion-related validity on the original study was calculated between the "language age" score obtained by the examiners (which was measured through SCID - Sequenced Inventory of Communication Development - Hedrick et al., 1975) and the total score of SVBRF. Pearson correlation was calculated as r = .81, p < .01, n = 83. Also, the total of the Expressive Language Communication Age

score of SCID and the total raw scores of First Use and Communicative of SVBRF were compared (r = .82 p < .01, n = 27).

Construct-Identification Validity

Factor Analysis

Although it was not performed on the original study, an exploratory factor analysis was run in this study. Each item of the analysis regarding the language sample included in the SVBRF-TV was considered as a scale item in order to test the Construct-Identification Validity of the test (N=169). Before conducting the factor analysis, the KaiserMeyer-Olkin (KMO) coefficient and Barlett test were calculated to evaluate whether the data were appropriate to conduct a factor analysis.

The KMO value for the factor analysis of 8 items was 0.72. According to the literature, the minimum KMO value should be above .60, and the Barlett test result should be significant in order to run a factor analysis on the data (Büyüköztürk, 2007; Pallant, 2001).

Result of the KMO value appears to be above the required threshold value, and the result of the Barlett test for 8 items were calculated as (x2=2568,399; df=28; p<.001). In order to determine the construct validity of the scale, the factors, whose eigenvalues were above 1.00 according to Kaiser normalisation, were considered as a measure. At the result of the analysis, only one factor was obtained that explained the 79,781% of the total variance. It appears that the factor loading values of the items varied between 0.83 and 0.93.

Correlation

On the level of construct-identification validity, correlations among the *repetitive, noncommunicative, nonintelligible* and *babbling* scores, used for the calculation of autistic speech score, were calculated with the use of Pearson correlation coefficient. Inter-item correlation coefficient constituting the autistic speech score of SVBRF was displayed on Table 2. The inter-item correlation coefficients reported on the original study was displayed in Italics on the table.

Table 2.

The Table of Inter-Item Correlation Coefficients for the Items Constituting The Autistic Speech Score on SVBRF-TV

	Repetitive	Noncommunicative	Nonintelligible	Babbling
Repetitive	1.00			
Noncommunicative	.76*/*62*	1.00		
Nonintelligible	.72*/*52*	.84*/*79*	1.00	
Babbling	.70*/*61*	.75*/*81*	.75*/*76*	1.00

*p < .01

Although it was reported on the original study, the degree of distinguishing children with ASD from non-ASD was examined through independent t-test samples. The result of t-test was calculated as $t_{(159)}$ = 19.2, p < .01. According to the test results, the difference between the mean score for autistic speech, which children with ASD got from SVBRF-TV, and the mean score of non-ASD is statistically noteworthy.

Findings on Reliability-Validity Regarding the Scores Obtained from IARF-TV

On the original study, Kuder-Richardson Formula for item reliability method was used for interrater reliability coefficient in order to examine the reliability of IARF. On the level of validity, t values and the levels of significance were determined following the analysis of IARF's sample profile in order to identify intergroup alteration on the level of construct-identification validity.

Also, differentiations on the sublevels among the groups of children with ASD and

non-ASD children, t values and significance levels were identified on the level of construct-identification validity. The reliability and validity of the Turkish version of the tool were examined with the use of same methods on this study, and also the correlations among defined behavioural categories were reported.

Interrater reliability

The IARF-TV tool includes 48 monitorings. These monitorings are classified in 12 subcells according to the student's behaviour and the intervals at which the monitoring was performed. The cells where the classification was done were compared through two observers. If the number of monitoring and location of the cell are same for both observers, that item was scored with "1", while the same monitoring was scored with "0" in the event of the number of monitoring was classified in different cells by both observers. A scale of 48 items was obtained through this scoring table, and the interrater reliability was calculated with KR-21 Formula. In terms of reliability, De Vellis (2014) states that the Kuder-Richardson coefficient is equal to alpha if the items forming a scale have two values. Büyüköztürk (2010) states that the KR-20 reliability coefficient can be used when the answers for the test items have two choices, but KR-21 can be used when the levels of difficulty of the test items do not differ greatly. Since the difficulty levels of the items included in the scale, which was obtained in this study and had 48 items and two values (1 and 0), and also since the same coefficient was used in the original scale, the inter-rater reliability was calculated by using the KR-21 formula. The interrater reliability coefficient, calculated with KR-21 formula at the end of the assessment by two observers on the basis of 32 subjects, was found as 0.83 (r_{32}=.83). The KR-21 reliability coefficient reported on the original study was 0.86.

Construct-Identification Validity

On the level of Construct-Identification Validity of IARF-TV, the degree of distinguishing children with ASD from non-ASD according to the defined behavioural categories was examined through independent samples t-test analysis. The results of the analysis were displayed on the Table 3.

Table 3.

Defined Behavioural Categories t Values Table on IARF-TV

Behavioural Category		Sampling Assessment $\overline{X}$	s	sd	*t* value
Interaction	ASD	13.90	8.33	88	4.76*
	Non-ASD	23.09	9.98		
Constructive Independent Play	ASD	11.08	8.79	88	1.65
	Non-ASD	14.62	11.46		
Unresponsiveness	ASD	21.13	15.35	88	4.05*
	Non-ASD	9.93	9.84		
Negative Aggressive	ASD	3.04	9.69	88	1.80
	Non-ASD	0.33	0.93		

* p < .01

There is a significant difference between the mean scores of children with ASD and non-ASD on the interaction and unresponsiveness levels according to the results of the analysis. On the levels of constructive independent play and negative aggressive, no significant difference was observed between the mean scores of children with ASD and non-ASD. The findings show that children with ASD could significantly be distinguished from the non-ASD on the level of interaction and unresponsiveness of IARF-TV, however, no significant difference was seen on the levels of

constructive independent play and negative aggressive categories.

There is a significant correlation between the categories of unresponsiveness and interaction in a negative way on the correlation analysis performed in order to examine the relationships among the four behavioural categories defined as parallel to these findings r = -.66 p < .01. Similarly, a significant correlation was observed in a negative way between the categories of unresponsiveness and constructive independent play r = -.58 p < .01.

Considering the diagnostic criteria for ASD, restrictions on social interaction are seen as the most significant characteristics of ASD which distinguish ASD from the other disability groups. Findings show that the tool can distinguish children with ASD from non-ASD on the basis of basic diagnostic criteria.

Discussion

Reliability

In order to examine reliability, first the method of two semi-test is used. As this method requires a single test form, a single student group and a single test execution, it was pointed out that it is widely used for examining the reliability for tests (Özen, Gülaçtı, Kandemir, 2006). The reliability of two semi-tests is the correlation coefficient calculated for the entire test with the use of Spearmen-Brown formula based on the relationship between the two halves of the test after the test items are divided into two equal halves. The reliability coefficient for the half of the test was calculated as .86, and for the entire test as .92 with the use of Spearman-Brown Formula on the reliability analysis of the two semi-tests for the SVBRF-TV tool. The resulting figures, which were reported on the original study, were respectively .95 and .97. Büyüköztürk (2010) points out that the fact that reliability coefficient calculated for a psychological test is .70 and above will generally be sufficient for the reliability of that test.

Şencan (2005) states that the high reliability coefficient rule can be made flexible to some extent only to do a research and develop a certain scale, and .70 can be accepted as a low level for reliability coefficients on the social science researches. On this study, the value calculated for SVBRF-TV is significant as a high value and statistically for the test reliability. Temel, Ersoy, Avcı and Turla (2004) reported the coefficient of test reliability of the two semi-tests on the Gazi Early Development Assessment Tool study.

Similarly, on another study conducted with the use of Spearman-Brown formula, Yılmaz-Irmak, Tekinsav-Sütçü, Aydın and Sorias (2007) reported a reliability result between .61 and .84 for two semi-test reliability examined on the basis of sub-levels for the study of the Turkish Adaptation of Autism Behavioural Check List. Although the value of .70 was found as a lower limit, the reliability coefficients calculated for the entire scale, and statistical significance were also discussed as supportive findings. Also the fact that there are multiple analysis findings instead of a single finding for the reliability and validity of a tool can be considered as objective proof that strengthens reliability and validity. It could be claimed that the test reliability coefficient calculated for the SBRF-TV tool on this study provides the level of acceptance for the reliability of the scores obtained from the tool.

The second statistical analysis study conducted in order to examine reliability is test-retest reliability study. Test-retest reliability is explained with the correlation among the scores obtained by performing a test twice on the same group at certain intervals. Test-retest reliability is used in order to interpret the extent to which a test performs definitive measurement (Büyüköztürk, 2010). On the study of identifying psychometric characteristics, Pearson correlation coefficients calculated on the basis of the data obtained from 33 participants at two weeks intervals for the SVBRF-TV tool vary among four sub-dimensions constituting the autistic speech score and three sub-dimensions constituting total score and total speech category and for the total score between .82 and .96. The test-retest correlation values reported on the original study are between .81 and .95.

It was reported that the test-retest reliability coefficient should be .80 and above (Innes and Straker, 2003). However, there are sources indicating the fact that test-retest

reliability or reliability coefficient of .70 can be acceptable (Büyüköztürk, 2010; Şencan, 2005). The values reported during the adaptation of the tool into Turkish were test-retest reliability correlations between .72 and .99 (Diken, Ardıç, Diken and Gilliam, 2012; Alev, 2011; Kapçı, Küçüker and Uslu, 2010; Yalaz, Anlar and Bayaoğlu, 2010; Köse, Bora, Erermiş and Aydın, 2010; Korkmaz, 2009; İncekaş, 2009; Atamaz, Yağız On, Durmaz, 2007; Maviş, Colay, Topbaş, Tanrıdağ, 2007; Düver, 2006; Colay, 2006; Savaşır, Sezgin and Erol, 2006; March-Göçer, 1996).

In the light of this information, it could be said that test-retest reliability coefficient calculated for SVBRF-TV is adequate for the tool to be considered reliable. Interexaminers or interrater reliability coefficient is a sub-type of equivalence analyses which are from the types of reliability analyses. Similar to parallel forms reliability, it is based on the fact that two examiners/raters score the same performance in order to measure the same characteristics (Büyüköztürk, 2010; Şencan, 2005). The degree of the relationship between two assessments or observations could be reported with overlap percentages or Pearson correlation coefficients. Interrater reliability on special needs education is generally considered to be reported with percentile type and suggested reliability is required to be close to hundred percent (Erbaş, 2012). One other analysis used on this study for interrater reliability is Kuder Richardson reliability. This coefficient is also used in order to examine the internal consistency among the test scores. KR-20 coefficient is used in cases where the questions for test items require only two response alternatives such as yes/no or true/false. In cases where the difficulty levels of test items do not differ considerably, KR-21 coefficient is used (Büyüköztürk, 2010). On the original study and mechanisms where raters monitor the same case on this study, this analysis was performed with the use of data obtained by coding of true/yes (1) for the consensus of two raters and false/no (0) for dissensus of two rates.

On the study of identifying the psychometric characteristics of SVBRF-TV, interrater reliability was obtained with the researcher monitoring and scoring the same performance with the examiner on 16 applications. The consensuses between two scorings were accepted as the reliability percentage of the tool. The mean of the percentage values calculated for three sub-levels of the tool was taken as the total value. The value calculated in the mean is 89%. On the original study, the value of 90% was reported. Kuder Richardson reliability for IARTF-TV was calculated as .83. Kuder Richardson reliability coefficient reported on the original study is .86.

Günayer Şenel (1998) used KR-21 reliability coefficient on the adaptation of Bangor Dyslexia Test into Turkish, and reported this value as .75. During the adaptation of scale development into Turkish, the compatibility rate was reported between 76% and 95%, and correlation coefficients between .27 and .99 (Kapçı, Küçüker and Uslu, 2010; Yalaz, Anlar and Bayoğlu, 2010; İncekaş, 2009; Yılmaz-Irmak, et al., 2008; Maviş, Colay, Topbaş, and Tanrıdağ, 2007; Düver, 2006; Colay, 2006; Akgün, 2005; Ege, Acarlar ve Turan, 2005; Temel, Ersoy, Avcı and Turla 2004; March-Göçer, 1996).

In the light of this information, it could be said that inter-examiners/raters reliability coefficients and rates calculated for SVBRF-TV and IARF-TV are adequate for the tool to be considered reliable. Item-total score correlation explains the relationship between the score obtained from test items and total score of the test. The fact that item-total correlation is positive and high shows that the items sample similar behaviours, and the internal consistency of the test is high. It was stated that the Cronbach Alpha coefficient that will be calculated for scales should be above .70, and the item-total correlation coefficient of the items should not be below .30. However, it was indicated that this value might be low on scales containing fewer items (less than 10 items), where the number of items on a scale affects the Cronbach Alpha coefficient, and therefore, it would be appropriate to report values for inter-item correlation coefficients (Büyüköztürk, 2010; Pallant, 2001).

Although it was not examined on the original study of the scale, item-total correlation was examined on this study since it provides empirical evidence regarding the

internal consistency of the scale, widely reported in many similar studies and provides a global reliability coefficient. Cronbach Alpha coefficient calculated for the internal consistency of SVBRF-TV is α = .92. Item-total correlation of each item varies between .78 and .87. The values obtained for SVBRT-TV provide reliability as values above the expected .70 and .30. Although these subvalues were offered, Cronbach Alpha and item-total correlation coefficients between .32 and .99 were reported on the adaptation of scale development into Turkish (Alev, 2011; Kapçı, Küçüker and Uslu, 2010; Köse, Bora, Erermiş and Aydın, 2010; Maviş, Colay, Topbaş, and Tanrıdağ, 2007; Colay, 2006; Savaşır, Sezgin and Erol, 2006; Ege, Acarlar and Turan, 2005; Kabil, 2005). In the light of this information, it could be said that test-retest reliability coefficient calculated for SVBRF-TV is adequate for the tool to be considered reliable.

Validity

The first method used for examining the validity of the tools, whose psychometric properties were identified, is the method of Criterion-Related Validity. Criterion-Related Validity is a validity technique that examines the relationship between test scores and one or a few external criteria. This technique of validity is divided into two as concurrent validity and predictive validity. On concurrent validity, which was also used on this study, the correlation between the scores obtained from a test to be improved, and from an old test measuring the same behaviour or another related test measuring another property was examined (Büyüköztürk, 2010).

Şencan (2005) points out that validity coefficients are not as high as reliability coefficients. According to Şencan, the "*r*" value in relation to validity and prediction varied between .30 and .50 and rarely went above .69. Thus, if a value in relation to validity and definiteness is between .30 and .50, it is concluded that the test is valid.

For the criterion-related comparisons, correlations between the "same" and "related" structures are examined. While the correlation values obtained through tests measuring the same conceptual structure are expected to be strong values such as .70-.80, it was stated that the values between .50 and .70, which show the medium relationships on the correlations made with the tests measuring related conceptual structures, could be evaluated as validity proof. However, Büyüköztürk (2010) indicates that the correlation coefficients above .30 could be used as validity proof for the test although they vary depending on the required specification. The validity of SVBRF-TV was examined through the method of criterion-related validity. The correlation coefficient value between SVBRF-TV total speech score and TEDIL receptive language score is .71, and the correlation coefficients among the subscores of 2/GARS-2-TV are .40 and .70. The correlation coefficients reported between SVBRF-TV and one other language tool are .81 and .82 on the original study. Criterion-related validity findings between .40 and .91 were reported on similar scale development studies, (Alev, 2011; Korkmaz, 2009; İncekaş, 2009; Savaşır, Sezgin and Erol, 2006). When the correlation findings between the tools of SVBRF-TV and TEDIL with 2/GARS-2-TV are examined, it appears that the value obtained through TEDIL is between the expected range, and that the tool provides proof for its validity. As for findings obtained through the 2/GARS-2-TV tool, it appears that the correlation between the autistic speech score of SVBRF-TV and social interaction subscale of 2/GARS-2-TV is .49. If the expected correlation obtained through a tool that measures a related structure to be between .50 and .70, the difference of .01 may be due to a measuring error. Again, considering that this comparison is made with the autistic speech scores obtained from SVBRF-TV and social interaction sub-scale of 2/GARS-2-TV, the fact that child's autistic speech level forms a strong correlation with the level of social interaction reported by teachers is logical.

Yet there are both child's observed momentary performance scores and scores created by the teacher's statements. Also, when the sources stating that the correlation coefficients above .30 provide evidence for the validity of the test are taken into account, it can be seen that the obtained correlation coefficients provide evidence for the validity. In accordance with this information,

it can be said that the correlation coefficients obtained for criterion related validity between SVBRF-TV and 2/GARS-2-TV tools are sufficient enough for the tools to be deemed valid.

The second method used for examining the validity of the tools is the construct-identification validity method. Construction-identification shows the degree of assessing an abstract concept (factor) accurately as part of the behaviour aimed to assess. It was stated that the techniques of factor analysis, cluster analysis, internal consistency analysis and hypothesis testing could be used for examining construction-identification. Hypothesis test technique also refers to concurrent validity; i.e., studies on construct-identification validity as part of concurrent validity provide evidence for construct-identification at the same time. Again, the level of distinguishing the groups with various characteristics is assessed under the same title (Büyüköztürk, 2010). Şencan (2005) states that construct-identification validity can be analysed through an analysis of inter-group differences. For instance, the fact that the ensemble average of the total scores or scores from the sublevels that the children with ASD or non-ASD groups obtained are significantly different could be evidential for the construct-identification validity.

Again, it is pointed out that in the event that the level of interrelationships of sub-dimensions, which form the same structure, provides evidence for construct-identification validity, and sublevels have the coefficient value of .60 and above, it could be claimed that levels are connected with each other and measure the same conceptual structure. It is also stated that Cronbach's Alpha values provide evidence for construct-identification validity and could be used for validity analysis. On the study, conducted for identifying psychometric properties, intergroup differences were examined using t-test and the relationships among inter-sublevels with Pearson correlation coefficient. On the studies on inter-group differences conducted using t-test analysis for SVBRF-TV and IARF-TV, significant differences were observed among the ensemble averages. Therefore, in view of the above theoretical explanations on construct validity are considered, it appears that the above mentioned tools can identify children with ASD, and these findings provide evidence for the construct validity of the tools. Construct validity analyses, which were conducted with the use of examining inter-subtest correlations, were performed for SVBRF-TV. The correlations among the four sub-levels that constitute the total autistic speech score of the scale in SVBRF-TV are between .72 and .84. The values reported for the same sub-levels on the original study vary between .52 and .81. In view of the results of the analyses of intergroup differences, inter-sublevels correlation and their explanations, the findings obtained for the construct validity of the tools, whose psychometric properties were identified, display that the tools are adequate for them be considered valid.

References

Alev, G. (2011). *Pragmatik Dil Becerileri Envanteri'nin Türkçe standardizasyon çalışması* (Turkish Standardisation of Pragmatic Language Skills Inventory) (Published master's thesis, Anadolu University, Institute of Medical Sciences, Eskişehir, Turkey). http://tez.yok.gov.tr/UlusalTezMerkezi adresinden elde edilmiştir. (Tez no: 288468)

Allen, R. A., Robins, D. L., & Decker, S. L. (2008). Autism spectrum disorders: Neurobiology and current assessment practices. *Psychology in the Schools, 45*(10), 905-917.

American Psychiatric Association (APA): *Diagnostic and Statistical Manual of Mental Disorders, Fourth Revised* Edition (DSM-IV-TR), American Psychiatric Association, Washington DC, Translated by Köroğlu, E., Hekimler Yayın Birliği, Ankara 2001.

Atamaz, F., Yağız On, A. ve Durmaz, B. (2007). EGE Aphasia Test: Description of the test and performance in normal subjects. *Turkish Journal of Physical Medicine Rehabilitation, 53*(1), 5-10.

Avcıoğlu, H. (2011). *Özel eğitime gereksinim duyan öğrencilerin eğitsel ve davranışsal değerlendirilmesi.* Ankara: Vize Yayıncılık.

Büyüköztürk, Ş. (2010). *Sosyal bilimler için veri analizi el kitabı* (12. Baskı). Ankara: Pegem Akademi.

CDC, (2016). *Center for Disease Control and Prevention*. http://www.cdc.gov

Colay, İ. K. (2006). *Gülhane Afazi Testi (GAT-2)'nin standardizasyon ve geçerlik ve güvenirlik çalışması* (Yayınlanmış yüksek lisans tezi, Anadolu Üniversitesi, Sağlık Bilimleri Enstitüsü, Eskişehir, Türkiye). http://tez.yok.gov.tr/UlusalTezMerkezi adresinden elde edilmiştir. (Tez no: 193095)

De Vellis R. (2014). *Scale development.* (*Ölçek geliştirme kuram ve uygulamalar)* Translated by Totan T., Ankara: Nobel Akademi Yayıncılık.

Diken, İ. H., Ardıç, A., Diken, Ö. ve Gilliam, J. E. (2012). Exploring the validity and reliability of Turkish Version of Gilliam Autism Rating Scale-2: Turkish standardization study. *Eğitim ve Bilim, 166*(37), 318-328.

Dover, C. J., & Le Couteur, A. (2007). How to diagnose autism. *Archives of Disease in Childhood, 92*, 540-545.

Düver, E. (2006). *5-7 yaş grubu normal gelişim gösteren ve özel gereksinimi olan çocukların dil kullanım (pragmatik) becerilerinin karşılaştırılması*. Yayınlanmamış Yüksek Lisans Tezi. Hacettepe Üniversitesi, Ankara.

Ege, P., Acarlar, F. ve Turan, F. (2005). *Ankara Artikülasyon Testi El Kitabı*. Ankara: Ankara Üniversitesi Yayınları.

Erbaş, D. (2012). Güvenirlik. Elif Tekin-İftar (Ed.) *Eğitim ve Davranış Bilimlerinde Tek Denekli Araştırmalar* içinde Ankara: TPD yayınları.

Gay, L. R., Mills, G. E., & Airasian, P. (2006). *Educational research: Competencies for analysis and application* (8th ed.). New Jersey: Prentice Hall.

Glosser, R. (2011*). Examination of the relationship between the child behavior checklist/6-18 and the social responsiveness scale parent forms using individuals with high functioning autism*. (Doctoral Dissertation). Retrieved from https://dspace.iup.edu/handle/2069/352?show=full

Günayer Şenel, H. (1998). *Okuma güçlüğü olan ve olmayan ilkokul öğrencilerinin okuma yazma düzeylerinin ve dislektik özelliklerinin karşılaştırılması.* (Yayınlanmamış doktora tezi). Ankara Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.

Gürsel, O. (2012). Özel eğitimde değerlendirme. İbrahim H. Diken (Ed.) *Özel eğitime gereksinimi olan öğrenciler ve özel eğitim* içinde. (5. Basım). Ankara: PEGEM Akademi.

Güven, O. S. (2009). *Erken dil gelişimi testi-Üçüncü edisyon'un Türkçeye uyarlama güvenirlik ve geçerlik ön çalışması* (Yayınlanmış yüksek lisans tezi, Anadolu Üniversitesi, Sağlık Bilimleri Enstitüsü, Eskişehir Türkiye). http://tez.yok.gov.tr/UlusalTezMerkezi adresinden elde edilmiştir. (Tez no: 236587)

Güven, O. S., & Topbaş, S.(2009). *Validity and reliability measures of TELD-3 to Turkish*. ASHA Convention, 18-21 Nov., New Orleans, US.

Hedrick, D., Prather, E. ve Tobin, A. (1975). *Sequenced inventory of communication development*. Seattle: University of Washington Press.

Hresko, W. P., Reid, D. K., & Hammill, D. D. (1999). *Test of Early Language Development: Teld 3: Examiner's Manual*. USA: Proed.

Innes, E., & Straker, L. (1999). Reliability of work-related assessments. *Work: A Journal of Prevention, Assessment and Rehabilitation, 13*(2), 107-124.

İncekaş, S. (2009). *Çocukluk Otizmini Derecelendirme Ölçeği geçerlik ve güvenilirlik çalışması* (Yayınlanmış tıpta uzmanlık tezi, Dokuz Eylül Üniversitesi, Tıp Fakültesi, İzmir, Türkiye) http://tez.yok.gov.tr/UlusalTezMerkezi adresinden elde edilmiştir. (Tez no: 230129)

Kabil, A. T. (2005). *A validity study of the Checklist for Autism in Toddlers (CHAT) on a Turkish sample*. (Yayınlanmamış yüksek lisans tezi). Boğaziçi Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul.

Kapçı, E. G., Küçüker, S. ve Uslu, S. (2010). How applicable is the Ages and Stages Questionnaires to Turkish Children?

Topics in Early Childhood Special Education, 30(3), 176-188.

Karasar, N. (2008). *Bilimsel araştırma yöntemi* (18. Baskı). Ankara: Nobel Yayın Dağıtım Tic. Ltd. Şti.

Korkmaz, E. (2009). *6-24 aylardaki normal gelişim gösteren ve gelişimsel gecikmesi olan bebeklerin iletişim ve sembolik davranışlarının karşılaştırılması*. (Yayınlanmamış yüksek lisans tezi). Hacettepe Üniversitesi, Sağlık Bilimleri Enstitüsü, Ankara.

Köse, S., Bora, E., Erermiş, S. ve Aydın, C. (2010). Otizm Spektrum Anketi Türkçe Formu'nun piskometrik özellikleri. *Anadolu Psikiyatri Dergisi, 11*, 253-260.

Krug, D. A., Arick, J. R. ve Almond, P. J. (2008). *ASIEP-3 Autism Screening Instrument for Educational Planning*. Third Edition. Austin, Texas: Pro-ed.

Maviş, İ., Colay, K., Topbaş, S. ve Tanrıdağ, O. (2007). Gülhane Afazi Testi (GAT-2)'nin standardizasyon ve geçerlik ve güvenirlik çalışması. *Türk Nöroloji Dergisi, 13*(2), 89-98.

March-Göçer, E. (1996). Frenchay Afazi Tarama Testi: Türk nöroloji hastaları için standardizasyon çalışması. *Türk Psikoloji Dergisi, 11*(38), 56-63.

Özen, Y., Gülaçtı, F. ve Kandemir, M. (2006). Eğitim bilimleri araştırmalarında geçerlik ve güvenirlik sorunsalı. *Erzincan Eğitim Fakültesi Dergisi, 8*(1), 69-89.

Pallant, J. (2001). *SPSS Survival Manual. A step bay step guide to data analysis using SPSS*. Philadelphia: Open University Press.

Peterson, N. L. & Meier, J. H. (1988). Assessment and evaluation processes. In Peterson, N. L. (Ed.). *Early intevention for handicapped and at risk children an introduction to early childhood-special education* (pp. 275-326). USA: Love Publishing Company.

Savaşır, I., Sezgin, N. ve Erol, N. (2006). *AGTE, Ankara Gelişim Tarama Envanteri*. Ankara: Türk Psikologlar Derneği.

Sucuoğlu, B. (2012). Otizm spektrum bozukluğu olan çocukların değerlendirilmesi. Elif Tekin İftar (Ed.). *Otizm spektrum bozukluğu olan çocuklar ve eğitimleri* içinde. Ankara: Vize Basın Yayın.

Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlik*. Ankara: Seçkin Yayıncılık.

Temel, F., Ersoy, Ö., Avcı, N. ve Turla, A. (2004). *Gazi Erken Gelişimi Değerlendirme Aracı "GECDA".* Ankara: Rekmay Ltd. Şti.

Tincani, M. (2010). Autism spectrum disorder. In Smith, D. D. & Tyler, N. C. (Eds.). *Introduction to special education* (pp.402-431). NJ: Pearson&Merilll.

TPD (Türk Psikologlar Derneği), (1997). *Eğitimde ve psikolojide ölçme standartları.* Amerikan Eğitim Araştırmaları Birliği, Amerikan Psikoloji Birliği, Eğitim Ölçümleri Uluslararası Konseyi Yayını. Türkçeye çevirenler: Hovardaoğlu, S. ve Sezgin, N. Ankara: Türk Psikologlar Derneği ve Öğrenci Seçme ve Yerleştirme Merkezi Yayını.

Yalaz, K., Anlar, B. ve Bayaoğlu, B. U: (2010). *Denver II Gelişimsel Tarama Testi Türkiye Standardizasyonu*. Ankara: Gelişimsel Çocuk Nörolojisi Derneği Yayınları.

Yılmaz-Irmak, T., Tekinsav-Sütçü, S., Aydın, A. ve Sorias, O. (2007). Otizm Davranış Kontrol Listesi'nin (ABC) geçerlik ve güvenirliğinin incelenmesi. *Çocuk ve Gençlik Ruh Sağlığı Dergisi, 14*(1), 13-23.