Reliability and Validity of the Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version, DSM-5 November 2016-Turkish Adaptation (K-SADS-PL-DSM-5-T)

Fatih ÜNAL[1], Ferhunde ÖKTEM[2], Füsun ÇETİN ÇUHADAROĞLU[3], S.Ebru ÇENGEL KÜLTÜR[4], Devrim AKDEMİR[5], Dilşad FOTO ÖZDEMİR[6], Halime Tuna ÇAK[7], Dilek ÜNAL[8], Kevser TIRAŞ[9], Cihan ASLAN[10], Bilge Merve KALAYCI[11], Büşra Sultan AYDOS[12], Funda KÜTÜK[13], Emine TAŞYÜREK[14], Remzi KARAOKUR[15], Başak KARABUCAK[16], Burak KARAKÖK[17], Yusuf KARAER[18], Abdülbaki ARTIK[19]

SUMMARY

Objective: The aim of this study was to evaluate the reliability and validity of the Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version, DSM-5 November 2016 -Turkish Adaptation (K-SADS-PL-DSM-5-T).

Method: A total of 150 children and adolescents between 6 and 17 years of age were assessed with K-SADS-PL-DSM-5-T. The degree of agreement between the DSM-5 criteria diagnoses and the K-SADS-PL-DSM-5-T diagnoses were considered as the measure of consensus validity. In addition, concurrent validity was examined by analyzing the correlation between the diagnoses on K-SADS-PL-DSM-5-T and relevant scales. Interrater reliabilities were assessed on randomly selected 20 participants. Likewise, randomly selected 20 other participants were interviewed with K-SADS-PL-DSM-5-T three weeks after the first interview to evaluate test-retest reliability.

Results: The consistency of diagnoses was almost perfect for eating disorders, selective mutism and autism spectrum disorder (κ=0.92-1.0), substantial for elimination disorders, obsessive-compulsive disorder, oppositional defiant disorder, generalized anxiety disorder, social anxiety disorder, depressive disorders, disruptive mood dysregulation disorder and attention deficit hyperactivity disorder (κ=0.67-0.80). Interrater reliability was perfect for selective mutism (κ=1.0), substantial for oppositional defiant disorder, disruptive mood dysregulation disorder, attention deficit hyperactivity disorder, depressive disorders and social anxiety disorder (κ=0.63-0.73). Test-retest reliability was almost perfect for autism spectrum disorder (κ=0.82), substantial for attention deficit hyperactivity disorder, oppositional defiant disorder, disruptive mood dysregulation disorder, depressive disorders and generalized anxiety disorder (κ=0.62-0.78).

Conclusion: The results of this study show that the K-SADS-PL-DSM-5-T is an effective instrument for diagnosing major childhood psychiatric disorders including selective mutism, disruptive mood dysregulation disorder and autism spectrum disorder which have recently been added to the schedule.

Keywords: K-SADS-PL, validity, reliability.

Received: 04.12.2018- **Accepted:** 25.12.2018

[1, 3, 4]Prof., [2]Ph.D. Prof., [5, 6, 7]Assoc. Prof., [8]Assis. Prof., [9, 13, 17, 18, 19]M.D., Hacettepe University, Faculty of Medicine, Department of Child and Adolescent Psychiatry, Ankara, [10]M.D., Kırıkkale High Specialization Hospital, Kırıkkale, [11]M.D., Yenimahalle Training and Research Hospital, Department of Child and Adolescent Psychiatry, Ankara, [12]M.D., Dr. Sami Ulus Children's Hospital, Ankara, [14]M.D., Afyonkarahisar State Hospital, Afyon, [15]M.D., Dr. Nafiz Körez Sincan State Hospital, Ankara, [16]M.D., Karaman State Hospital, Karaman, Turkey.

e-mail: *dr.fatih.unal@gmail.com*

https://doi.org/10.5080/u23408

INTRODUCTION

One of the important difficulties in child and adolescent psychiatry is evaluating psychiatric symptoms in children and adolescents for a valid diagnosis. Contrary to many medical disorders, there is insufficient evidence, based on biological measurements for confirming the psychiatric diagnosis (Lempp et al. 2012). Clinical evaluation process becomes even more difficult due to the time and situation dependent inconsistencies in psychiatric symptoms, uncertainties in children's expressions, the changes in the mental state between the interviews, variability in information or biases of interviewers in rating the importance of symptoms (Duncan et al. 2018).

It has been suggested that structured clinical interviews may be useful to overcome these difficulties. Such interviews have some differences in terms of style and content from the usual clinical evaluation. Following the same interview line and content increases the reliability of the diagnosis or diagnoses made. With the help of the structured interviews, the onset, severity, and duration of the disorder, the temporal sequence and the degree of functional impairment can be evaluated in more detail. For these reasons, structured interviews can also be seen as a reliable guide for inexperienced interviewers (Leffler et al. 2015). Although it takes a little longer than the usual clinical psychiatric interview, it has been suggested that making more accurate diagnoses with the help of structured interviews could increase the effectiveness of treatment and, thereby, save time (Garcia-Barrera and Moore 2013).

Although there are many structured or semi-structured interview schedules used on children and adolescents, the most commonly used interview schedule is considered to be Schedule for Affective Disorders and Schizophrenia for School-Age Children (Ambrosini 2000, Boyle et al. 2017). The origins of this interview schedule are based on Schedule for Affective Disorders and Schizophrenia-SADS, developed by Endicott and Spitzer (1978) for adults and adapted by Chambers et al. (1985) for children.

The DSM-III and DSM-IV Present and Lifetime Version of the schedule was updated by Kaufman et al. (2016) after the DSM system was revised in 2013 (APA 2013). The previous DSM-III and DSM-IV version of the schedule was translated into Turkish and it was shown that it provided valid and reliable data for many diagnoses (Gokler et al. 2004). However, there is no validity and reliability study for the Turkish DSM-5 version of the schedule, which includes both dimensional and categorical diagnostic evaluations. This not only leads to problems regarding standardization of clinical evaluation but also has a negative effect on the quantity and quality of the research in the field of child and adolescent psychiatry in our country.

In this study, it is aimed to evaluate the validity and reliability of the Schedule for Affective Disorders and Schizophrenia for School-Age Children - Present and Lifetime Version, November 2016 – Turkish adaptation (K-SADS-PL-DSM-5-T), designed to evaluate 23 different diagnostic areas with the addition of disruptive mood dysregulation disorder, selective mutism, binge eating disorder, and autism spectrum disorder.

METHOD

Participants

The study group consisted of 120 children and adolescents aged between 6 and 17 years admitted to the outpatient clinic of the Hacettepe University Child and Adolescent Psychiatry Department for the first time between February 2018 and June 2018. Due to age limitation and/or low prevalence ratios, "first-time application criteria" was not applied to children with autism spectrum disorders, selective mutism, and disruptive mood dysregulation disorder. Children with any chronic medical problems or intellectual disabilities that could interfere with the evaluations were also excluded. The control group consisted of 30 children and adolescents in the same age range, admitted to the pediatric outpatient clinic of the same hospital. Psychiatric complaints or clinically evaluated intellectual disabilities or chronic medical problems that might interfere with the evaluations were the exclusion criteria for the control group.

The study was approved by Hacettepe University Clinical Research Ethics Committee (GO 18/282 - 05).

Assessment Tools

Schedule for Affective Disorders and Schizophrenia for School-Age Children - Present and Lifetime Version, K-SADS-PL-DSM-5 November 2016: Adaptation to the Turkish language of this semi-structured questionnaire, which Kaufman et al. (2016) updated according to DSM-5 diagnostic criteria, was carried out by the research team. K-SADS-PL-DSM-5-T took the final form after the review of the discrepancies between translation and back-translation of the original text, and the effectivity assessment of the schedule through pilot interviews with the parents and children.

In the first part of the schedule, an unstructured interview with the child and his/her family about the sociodemographic characteristics, complaints, development history, and health status, general information about the functionality in school and at home are questioned. In the second part, there are screening questions evaluating more than 200 symptoms, and the third part consists of the evaluation and observation results to confirm the DSM-5 diagnoses. Information from each section is scored separately first, but the final decision is based on the clinician's observations.

Primary diagnoses which have been updated according to DSM-5 diagnostic criteria that can be screened by K-SADS-PL-DSM-5-T, are major depression, persistent depression, mania, hypomania, cyclothymia, bipolar disorders, disruptive mood dysregulation disorder, schizoaffective disorders, schizophrenia, schizophreniform disorder, brief psychotic disorder, panic disorder, agoraphobia, separation anxiety disorder, simple phobia, social anxiety disorder, selective mutism, generalized anxiety, obsessive-compulsive disorder, attention deficit hyperactivity disorder, conduct disorder, oppositional defiant disorder, enuresis, encopresis, anorexia nervosa, bulimia, binge eating disorder, transient tic disorder, Tourette's disorder, chronic motor or vocal tic disorder, alcohol use disorder, substance use disorder, post-traumatic stress disorder, adjustment disorders, and autism spectrum disorder. Disruptive mood dysregulation disorder, selective mutism, and autism spectrum disorder newly introduced into DSM-5 were also included with the diagnoses that can be screened with K-SADS-PL-DSM-5-T (Kaufman et al. 2016).

Conners' Rating Scales-Revised (CRS-R): Evaluating the symptoms and severity of ADHD, this scale was developed by Connors (1997) for determination of 27 items that gave the highest factor load as a result of the factor analysis applied to the data collected for the earlier long forms. CRS-R consists of three subscales (oppositional, cognitive problems/inattention, and hyperactivity) and an auxiliary part (ADHD index) (Kaner et al. 2006).

Screen for Child Anxiety and Related Disorders (SCARED): The scale was developed by Birmaher et al. (1999) for screening childhood anxiety disorders. It includes subscales evaluating symptoms of somatization, panic, generalized anxiety, separation anxiety, social phobia and fear of school (Çakmakçı 2004).

Beck Depression Inventory (BDI): The scale is composed of 21 items evaluating the symptoms of depression in vegetative, emotional, cognitive and motivational areas. It was developed by Beck et al. (1961, 1988) and adapted by Hisli (1989) to the Turkish population.

Affective Reactivity Index (ARI): The scale, developed by Stringaris et al. (2012) evaluates irritability symptoms in the last six months with the help of child and parent forms. It can distinguish severe mood disorders from both healthy children and patients with bipolar disorder. Turkish validity reliability study was performed by Kocael (2015).

Childhood Autism Rating Scale (CARS): The scale is used to diagnose autism based on data obtained from the family interview and observation of the child and it consists of 15 items (Schopler et al. 1980). Sucuoğlu et al. (1996) conducted the validity and reliability study of the Turkish language version. The cut-off point of the Turkish language version was found to be 30 in a study conducted in a larger sample (Incekaş 2009).

Autism Behavior Checklist (AuBC): This 57-item assessment tool, consisting of five sub-scales, including sensory behavior, social relating, body and object use, language and communication skills, and social and adaptive skills, was developed by Krug et al. (1980). The scores of the scale can vary between 0 and 159, and the children diagnosed with autism spectrum disorder usually receive a score of 68 or higher (Irmak et al. 2007).

Selective Mutism Questionnaire (SMQ): The scale, which was developed by Bergman et al. (2008) in order to assess the level of inhibition of the child's speech in various situations, contains 17 statements describing typical situations in which talking is expected. In the present study, SMQ is the only scale in which low scores indicate more impairment. The Turkish version of SMQ, which was adapted by the researchers, has a strong internal consistency reliability coefficient (Cronbach α = .869).

Procedure

The researchers were clinically qualified children and adolescent psychiatrists, who had substantial training in structured interviewing and experience on using the K-SADS-PL. Children and adolescents who met the inclusion criteria were first evaluated with a clinical psychiatric interview based on the DSM-5 diagnostic criteria. Then the K-SADS-PL-DSM-5-T was applied by a different researcher. The percentage of agreement between these two evaluations was considered as the consensual validity. The concurrent validity was measured as the relation between the current diagnoses made with the K-SADS-PL-DSM-5-T and the scores obtained from the assessment scales associated with that diagnosis. Assessment scales were used on the age groups with a higher prevalence of the associated disorder. Thus, BDI evaluating depressive disorders is used only on adolescents, whereas the CARS and AuBC evaluating autism spectrum disorder, and the SMQ evaluating selective mutism are used only on children. On the other hand, the CRS-R evaluating attention deficit hyperactivity disorder, the SCARED evaluating anxiety disorders, and the ARI evaluating disruptive mood dysregulation disorder was applied to all age groups because of the large age distribution of these disorders.

Interrater reliability was examined through a re-evaluation of the video recordings of the K-SADS-PL-DSM-5-T interview with randomly selected 20 participants (who gave consent to the video recording) by a different rater. Test-retest reliability was examined by another researcher through a re-assessment of randomly selected different 20 participants on average 21.1+4.9 days (12-30 days) after the first evaluation. The clinical information about the participants and the evaluations of the first interviewer were not shared with the researchers who assessed the video recordings or conducted the second interview.

Data Analysis

A computerized statistical program was used to analyze the data. Categorical variables were evaluated by the chi-square test. The Kolmogorov-Smirnov test for continuous variables indicated that age and all scale scores did not show normal distribution. The Mann-Whitney-U test was used to evaluate concurrent validity, and kappa statistics (Cohen 1960) were used in order to evaluate consensual validity, interrater reliability, and test-retest reliability. When interpreting the kappa coefficient, the following criteria proposed by Landis and Koch (1977) regarding the degree of consensus were taken into account: κ=0.81-1.00 (almost perfect agreement), κ=0.61-0.80 (substantial), κ=0.41-0.60 (moderate), κ=0.21-0.40 (fair), κ=0.00-0.20 (slight), κ < 0.00 (no agreement).The disorders with an insufficient number of patients to perform the kappa statistics (negative values in 95% confidence interval calculated for the kappa values) were excluded. For disorders with a confidence interval greater than 0.50, validity and reliability were evaluated with consensus ratios, where 80% or higher figures were considered as the "lowest acceptable consensus ratio" as recommended by McHugh (2012).

RESULTS

Participants consisted of 64 children and 86 adolescents with a median age of 13 years (Range: 6-17 years), 50.7% (n=76) males and 49.3% (n=74) females, living in families with a median level of 11-year education (Range: 5-15 years) and a total median annual household income of 48.000 TRY (Range: 18.000-120.000 ₺), which was above the average income in Turkey (Turkish Statistical Institute 2015). There were not any statistically significant differences between the study and control groups in terms of age (U=1604, p=.35), gender (X^2=0.007, p=.94), maternal and paternal education levels (U=1360, p=.26; U=1258, p=.37), and annual household income (U=1848, p=.75).

Diagnostic Profile

The median number of K-SADS-PL-DSM-5-T diagnoses was 2 (Range: 1-8) for current diagnoses and 1 (Range: 0-8) for past diagnoses. Most of the participants (62.6%, n=94) had multiple diagnoses, only 31 patients (20.7%) met the diagnostic criteria of a single DSM-5 disorder, and 25 (16.7%) did not meet any diagnostic criteria. The most frequently made K-SADS-PL-DSM-5-T diagnoses were attention deficit hyperactivity disorder (n = 61, 40.7%), depressive disorders (n = 36, 24.0%), and generalized anxiety disorder (n = 29, 19.3%) (Table 1).

Table 1. Consensual Validity of KSAD-PL-DSM-5-T Diagnoses (n=150)

KSAD-PL-DSM-5-T Diagnoses *	N (%)	Consensus	Kappa
Attention Deficit Hyperactivity Disorder	61 (40.7)	84.6 %	.67 **
Depressive Disorders[a]	36 (24.0)	89.3 %	.70 **
Generalized Anxiety Disorder	29 (19.3)	92.6 %	.74 **
Oppositional Defiant Disorder	27 (18.0)	93.3 %	.76 **
Social Anxiety Disorder	17 (11.3)	95.3 %	.71 **
Autism Spectrum Disorder	15 (10.0)	98.7 %	.92 **
Obsessive-Compulsive Disorder	15 (10.0)	96.7 %	.79 **
Disruptive Mood Dysregulation Disorder	14 (9.3)	96.0 %	.70 **
Elimination Disorders	12 (8.0)	97.3 %	.80 **
Eating Disorders[b]	11 (7.3)	100 %	1.0 **
Selective Mutism	10 (6.7)	99.3 %	.94 **

* Statistical analysis was performed only for current diagnoses with a sufficient number of patients. Due to comorbid diagnoses, the numbers in the table are more than the total number of patients.
**p<0.001
[a]Includes major depressive disorder and persistent depressive disorder (dysthymia). Disruptive mood dysregulation disorder is not included.
[b]Includes anorexia nervosa, bulimia nervosa and binge eating disorder.

Consensual Validity

Statistical analysis was performed only for 11 diagnostic groups to evaluate consensual validity, as other diagnoses had an insufficient number of patients. The rate of agreement between the clinical diagnoses based on DSM-5 criteria and the current K-SADS-PL-DSM-5-T diagnoses were statistically significant. Kappa values were found to be very good for eating disorders, selective mutism, and autism spectrum disorder; and good for elimination disorders, obsessive-compulsive disorder, oppositional defiant disorder, generalized anxiety disorder, social anxiety disorder, depressive disorders, disruptive mood dysregulation disorder, and attention deficit hyperactivity disorder (Table 1).

Concurrent Validity

Children and adolescents who met the diagnostic criteria for attention deficit hyperactivity disorder scored significantly higher than the ones without the diagnosis on the CRS-R. Children and adolescents who met the diagnostic criteria for panic disorder, agoraphobia, separation anxiety disorder, simple phobia, social anxiety disorder or generalized anxiety disorders had significantly higher scores on the SCARED compared to those without the diagnosis. Adolescents who met the diagnostic criteria for major depression or persistent depression scored significantly higher than the others on the BDI. Children and adolescents who met the diagnostic criteria for disruptive mood dysregulation disorder had significantly higher scores on the ARI compared to those without the diagnosis. Children who met the diagnostic criteria for autism spectrum disorder had significantly higher

Table 2. Concurrent Validity of KSAD-PL-DSM-5-T Diagnoses (n=150)

Attention Deficit Hyperactivity Disorder	**YES (n=52) Median Range**	**NO (n=57) Median Range**	**Statistics ***
Conners' Rating Scales-Revised	37 (16-48)	12 (2-33)	U=37 **
Anxiety Disorders[a]	**YES (n=38)**	**NO (n=56)**	**Statistics**
Screen for Child Anxiety and Related Disorders	39 (14-68)	14 (2-42)	U=167.5 **
Depressive Disorders[b]	**YES (n=32)**	**NO (n=38)**	**Statistics**
Beck Depression Inventory	38 (14-60)	9 (0-27)	U=47.5 **
Disruptive Mood Dysregulation Disorder	**YES (n=14)**	**NO (n=83)**	**Statistics**
Affective Reactivity Index – Self	8 (4-11)	3 (0-11)	U=142 **
Affective Reactivity Index – Parent	9 (6-12)	4 (0-11)	U=71 **
Autism Spectrum Disorder	**YES (n=15)**	**NO (n=43)**	**Statistics**
Autism Behavior Checklist	71 (52-96)	34 (12-57)	U=3.5 **
Childhood Autism Rating Scale	36 (28-42)	16 (15-32)	U=5.5 **
Selective Mutism	**YES (n=8)**	**NO (n=48)**	**Statistics**
Selective Mutism Questionnaire	18 (13-37)	41 (22-51)	U=14,5 **

*Mann Whitney U test
**p<0.001
[a]Includes generalized anxiety disorder, panic disorder, agoraphobia, separation anxiety disorder, social anxiety disorder, and simple phobia.
[b]Includes major depressive disorder and persistent depressive disorder (dysthymia).

scores on the CARS and AuBC compared to those without the diagnosis, and children who met the diagnostic criteria for selective mutism had significantly lower scores on the SMQ compared to those without the diagnosis. (Table 2).

Interrater Reliability

Interrater reliability was evaluated only for attention deficit hyperactivity disorder, depressive disorders, oppositional defiant disorder, disruptive mood dysregulation disorder, and selective mutism as there was an insufficient number of patients with the other diagnoses.

The percentage of interrater agreement on the diagnoses was found to be statistically significant in all of these diagnostic groups. Kappa values were found to be very good for selective mutism, and good for oppositional defiant disorder, disruptive mood dysregulation disorder, attention deficit hyperactivity disorder, and depressive disorders (Table 3).

Table 3. Interrater Reliability of KSAD-PL-DSM-5-T Diagnoses (n=20)

KSAD-PL-DSM-5-T Diagnoses*	**N**	**Consensus**	**Kappa**
Attention Deficit Hyperactivity Disorder	8	85 %	.68 **
Depressive Disorders[a]	7	85 %	.66 **
Oppositional Defiant Disorder	5	90 %	.73 ***
Selective Mutism	4	100 %	1.0 ***
Disruptive Mood Dysregulation Disorder	4	85 %	.63 **

*Statistical analysis was performed only for current diagnoses with a sufficient number of patients. Due to comorbid diagnoses, the numbers in the table are more than the total number of patients.
**p<0.05, ***p<0.001
[a]Includes major depressive disorder and persistent depressive disorder (dysthymia). Disruptive mood dysregulation disorder is not included.

Test-Retest Reliability

Test-retest reliability was evaluated only for 5 diagnostic groups due to an insufficient number of patients in the other diagnostic categories. The percentage of agreement between the diagnoses of the first and the second interview was statistically significant in all of these diagnostic groups. Kappa values were very good for autism spectrum disorder, and good for attention deficit hyperactivity disorder, oppositional defiant disorder, depressive disorders, and generalized anxiety (Table 4).

Table 4. Test-Retest Reliability of KSAD-PL-DSM-5-T Diagnoses (n=20)

KSAD-PL-DSM-5-T Diagnoses*	**N**	**Consensus**	**Kappa**
Attention Deficit Hyperactivity Disorder	7	90 %	.78 ***
Depressive Disorders[a]	6	85 %	.66 **
Oppositional Defiant Disorder	5	90 %	.73 **
Autism Spectrum Disorder	4	95 %	.82 ***
Generalized Anxiety disorder	4	85 %	.63 **

*Statistical analysis was performed only for current diagnoses with a sufficient number of patients. Due to comorbid diagnoses, the numbers in the table are more than the total number of patients.
**p<0.05, ***p<0.001
[a]Includes major depressive disorder and persistent depressive disorder (dysthymia). Disruptive mood dysregulation disorder is not included.

DISCUSSION

With the publication of the 5th edition of DSM, new definitions for many disorders in children and adolescents were made (Machado et al. 2013). Thus, updating the structured interview schedules became inevitable. Some researchers have focused on disorders that have just been

added or changed in the last edition such as autism spectrum disorder, social anxiety disorder, intermittent explosive disorder, disruptive mood dysregulation disorder, avoidant restrictive food intake disorder, and binge eating disorder (de la Peña et al. 2018a, 2018b). Kaufman et al. (2016) preferred preserving the previous format (Kaufman et al. 1997) and added most of these disorders to the schedule. Although data on the validity and reliability of this DSM-5-compatible clinician version has not yet been published, Kaufman et al. (2017) introduced a web-based computerized version of the schedule, and reported that this version had a strong convergent validity, and high inter-rater reliability for depressive disorders, anxiety disorders, attention deficit hyperactivity disorder, oppositional defiant disorder, and conduct disorder. However, the recently added disorders were not mentioned in this study. The fact that the current study includes data related to these newly added conditions or disorders with changed diagnostic criteria can be considered as a distinctive contribution to the literature.

In the current study, validity was evaluated by two different methods. The percentages of agreement between the clinical diagnoses based on DSM-5 criteria and the K-SADS-PL-DSM-5-T diagnoses were evaluated first. The agreement was found to be very good for eating disorders, selective mutism, and autism spectrum disorder (κ=0.92-1.00), and good for elimination disorders, obsessive-compulsive disorder, oppositional defiant disorder, generalized anxiety disorder, social anxiety disorder, depressive disorders, disruptive mood dysregulation disorder, and attention deficit hyperactivity disorder (κ=0.67-0.80).

There are four more studies evaluating the validity of the old version of the interview schedule with the same method. Shanee et al. (1997) who studied the schedule in an inpatient clinic, found the percentage of agreement between two interviews slightly higher than in the current study for attention deficit hyperactivity disorder, oppositional defiant disorder, and depressive disorders (κ=0.79-0.94). This difference in agreement ratios can be related to the fact that patients in an inpatient clinic might give more detailed information than outpatients. In another study (Shahrivar et al. 2010), also conducted in an inpatient clinic, where the majority of the patients were diagnosed with bipolar disorder, the percentage of agreement was slightly higher than in the current study for depressive disorders and attention deficit hyperactivity disorder (κ=0.74-0.80), but much lower for oppositional defiant disorder, generalized anxiety disorder, and obsessive-compulsive disorder (κ=0.17-0.48). This difference in agreement ratios can be explained by the fact that the assessment was made by relatively inexperienced clinicians (residents in child psychiatry rotation).In the study of Gokler et al. (2004), similarly to the current study, the agreement percentages were found to be good for attention deficit hyperactivity disorder and tic disorders (κ=0.64-0.72), but moderate for depressive disorders and anxiety disorders (κ=0.52-0.56). This difference in depressive disorders and anxiety disorders can be related to the relatively low number (n= 6, 11.5%) of adolescents in this study. In another study conducted in Korea in the same year (Kim et al. 2004), the agreement between the two interviews for attention deficit hyperactivity disorder was found to be similar to the result of the current study (κ=0.69). The researchers explained the lower agreement percentages for oppositional defiant disorder (κ=0.41), depressive disorders (κ=0.24), and anxiety disorders (κ=0.29), by the cultural characteristics of the Far East.

In a meta-analysis (Rettew et al. 2009) examining the agreement between clinical assessments and structured interviews as in the current study, the highest agreement was observed for anorexia nervosa (κ=0.86). Researchers observed that agreement between the two interviews on the other disorders was lower. They speculated that focusing on the complaints in clinical interviews could lead to associate symptoms with only one disease and prevent the interviewer from detecting comorbid disorders. Similarly, patients tend to give less information to the clinician in these interviews. On the other hand, some clinicians may be eager to diagnose a disorder associated with the main complaint even if the diagnostic criteria are not fully met, for instance, diagnosing ADHD with 5/9 symptoms in a child who has difficulty listening to the teacher.

In the current study, concurrent validity was also evaluated for some disorders by using self-report scales which were shown to be able to evaluate the severity of these disorders. Kaufman et al. (1997), who examined the validity of the K-SADS-PL with the same method, found the agreement for depressive disorders, anxiety disorders, attention deficit hyperactivity disorder, and conduct disorders statistically significant. Similarly, Brasil and Bordin (2010) showed that the CBCL scores significantly distinguished children diagnosed with the K-SADS-PL from non-diagnosed children. In the current study, the scales associated with disruptive mood dysregulation disorder, autism spectrum disorder, and selective mutism (in addition to attention deficit hyperactivity disorder, anxiety disorders, and depressive disorders) successfully distinguished those who met the diagnostic criteria for these disorders from those who did not.

In terms of interrater reliability, the agreement between the two interviewers was found to be very good for selective mutism (κ=1.0) and good for oppositional defiant disorder, disruptive mood dysregulation disorder, attention deficit hyperactivity disorder, and depressive disorders (κ=0.63-0.73). Similarly, Gökler et al. (2004) found that the agreement between the two raters was good for attention deficit hyperactivity disorder and depressive disorders (κ=0.63-0.68). Shanee et al. (1997) found this agreement to vary from very good to

good for depressive disorders and anxiety disorders (κ=0.64-1.0), Peña et al. (2018b) reported good agreement (>κ=0.7) with most of the disorders they evaluated, and Kaufman et al. (1997) reported significantly higher kappa values (κ=0.93-1.0) for inter-rater reliability.

When literature review is expanded for different versions of the schedule, kappa values ranging from 0.64-0.89 for K-SADS (Ambrosini et al. 1989) and 0.80-0.90 for K-SADS-P (Kolaitis et al. 2003) to 0.73-0.96 for K-SADS-E (Gau et al. 2005) were reported. The inconsistencies in the inter-rater reliability figures between the studies can be explained by differences in the methodology or the sampling of the study, dissimilarities in the number of evaluators, and the training received for the interview. The fact that the evaluators in the study (Kaufman et al. 1997) reporting the highest interrater reliability values consisted of a team working together in the process of developing the interview tool, supports this view.

In terms of test-retest reliability, the agreement between the two interviews was found to be very good for autism spectrum disorder (κ=0.82) and good for attention deficit hyperactivity disorder and oppositional defiant disorder, depressive disorders and generalized anxiety disorders (κ=0.63-0.78). The relatively low value from the DSM-5 field trials (κ=0.69) as compared to the current study, was the only data found in the literature for autism spectrum disorder (Regier et al. 2013). More data have been found on the other disorders. In terms of test-retest reliability, Kaufman et al. (1997) reported values similar to those in the current study for attention deficit hyperactivity disorder and oppositional defiant disorder (κ=0.63-0.74), but a slightly higher value for depressive disorders and anxiety disorders (κ=0.80-0.90). These values are similar to the kappa values reported by Gökler et al. (2004) for anxiety disorders (κ=0.78) and attention deficit hyperactivity disorder (κ= 0.89); by Kim et al. (2004) for attention deficit hyperactivity disorder (κ=0.75) and by Shahrivar et al. (2010) for oppositional defiant disorder (κ=0.77) and attention deficit hyperactivity disorder (κ=0.87). In a review of 31 studies (n= 3344) evaluating the test-retest reliability of structured interview scales used for children and adolescents, Duncan et al. (2018) reported that the pooled reliability was moderate (κ=0.58) and reliability was erroneously rated as significantly higher in studies with indicators of poor or fair study methodology quality.

The relatively large sample size, a balanced participation of children and adolescents, the presence of a control group, and the qualification of the interviewers are the strengths of the current study. The limitations of the study can be listed as (1) inclusion of children living in families with above the average annual income and education of Turkey, (2) lack of sufficient data on some disorders with low prevalence, (3) selection bias due to exceptions in "the first-time application criteria" for children with autism spectrum disorder, selective mutism, and disruptive mood dysregulation disorder, because of age limitation and low prevalence rates, (4) absence of psychometric information on measures for assessing disorders, which are rarely seen in certain age groups (e.g. BDI evaluating depressive disorders was not administered to children or SMQ evaluating selective mutism was not administered to adolescents), and (5) that the validity and reliability of the Turkish language version of the SMQ has not yet been published.

In addition, using video recording for interrater reliability can be seen as a limitation as it will not allow possible communications between the interviewer and the informant. However, many studies have shown that this method is as reliable as face-to-face interviews (Ambrosini et al. 1989, Birmaher et al. 2009, Chen et al. 2017). Finally, it should be kept in mind that reliability figures can be higher in clinical samples than in the community sample (Boyle et al. 1993), which limits the generalization of the findings.

CONCLUSION

The results of this study show that the K-SADS-PL-DSM-5-T is an effective instrument for diagnosing major childhood psychiatric disorders including selective mutism, disruptive mood dysregulation disorder and autism spectrum disorder which have recently been added to the schedule. Demonstration of the validity and reliability of this version for a large number of disorders, will improve both the quality of diagnosis and treatment in clinical settings, and increase the contribution of researchers to the literature in our country.

REFERENCES

Ambrosini PJ, Metz C, Prabucki K et al (1989) Videotape reliability of the third revised edition of the K-SADS. J Am Acad Child Adolesc Psychiatry 28: 723-8.

Ambrosini PJ (2000) Historical development and present status of the schedule for affective disorders and schizophrenia for school-age children (K-SADS). J Am Acad Child Adolesc Psychiatry 39: 49-58.

American Psychiatric Association (2013) Diagnostic and Statistical Manual of Mental Disorders. 5th Ed.(DSM-5). American Psychiatric Association, Washington DC.

Beck AT, Ward CH, Mendelson M et al (1961) An inventory for measuring depression. Arch Gen Psychiatry 4: 561-71.

Beck AT, Steer RA, Carbin MG (1988) Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. Clin Psychol Rev 8: 77-100.

Bergman RL, Keller ML, Piacentini J et al (2008) The development and psychometric properties of the selective mutism questionnaire. J Clin Child Adolesc Psychol 37: 456-64.

Birmaher B, Ehmann M, Axelson DA et al (2009) Schedule for affective disorders and schizophrenia for school-age children (K-SADS-PL) for the assessment of preschool children: a preliminary psychometric study. J Psychiatr Res 43: 680-6.

Birmaher B, Brent D, Chiapetta L et al (1999) Psychometric properties of

the Screen for Child Anxiety Related Emotional Disorders (SCARED): a replication study. J American Acad Child Adolesc Psychiatry 38: 1230-6.

Boyle MH, Offord DR, Racine Y et al (1993) Evaluation of the diagnostic interview for children and adolescents for use in general population samples. J Abnorm Child Psychol 21: 663-81.

Boyle MH, Duncan L, Georgiades K et al (2017) Classifying child and adolescent psychiatric disorder by problem checklists and standardized interviews. Int J Methods Psychiatr Res 26 (4). doi: 10.1002/mpr.1544.

Brasil HH, Bordin IA (2010) Convergent validity of K-SADS-PL by comparison with CBCL in a Portuguese speaking outpatient population. BMC Psychiatry 19: 83.

Chambers WJ, Puig-Antich J, Hirsch M et al (1985) The assessment of affective disorders in children and adolescents by semi-structured interview. Test-retest reliability of the schedule for affective disorders and schizophrenia for school-age children, present episode version. Arch Gen Psychiatry 42: 696-702.

Chen YL, Shen LJ, Gau SSF (2017) The Mandarin version of the kiddie-schedule for affective disorders and schizophrenia-epidemiological version for DSM-5-a psychometric study. J Formos Med Assoc 116: 671–8.

Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Meas 20: 37-46.

Conners CK (1997) Conners' rating scales-revised. Instruments for use with children and adolescents. Toronto, MHS.

Çakmakçı F (2004) Çocuklarda anksiyete bozukluklarını tarama ölçeği geçerlik ve güvenirlik çalışması. Yayınlanmamış Uzmanlık Tezi, Kocaeli, Kocaeli Üniversitesi Tıp Fakültesi.

Duncan L, Comeau J, Wang L et al (2018) Research Review: Test-retest reliability of standardized diagnostic interviews to assess child and adolescent psychiatric disorders: a systematic review and meta-analysis. J Child Psychol Psychiatry doi:10.1111/jcpp.12876.

Endicott J, Spitzer RL (1978) A diagnostic interview: the schedule for affective disorders and schizophrenia. Arch Gen Psychiatry 35: 837-44.

Garcia-Barrera MA, Moore R (2013) History-taking, clinical interviewing, and the mental status examination in child assessment. Oxford Handbook of Psychological Assessment of Children and Adolescents. DH Saklofske, CR Reynolds, VL Schwean (Eds), New York, Oxford University Press, p. 423-44.

Gau SSF, Chong MY, Chen THH et al (2005) A 3-year panel study of mental disorders among adolescents in Taiwan. Am J Psychiatry162: 1344-50.

Gökler B, Ünal F, Pehlivantürk B, Kültür EÇ et al (2004) Okul Çağı Çocukları için Duygulanım Bozuklukları ve Şizofreni Görüşme Çizelgesi-Şimdi ve Yaşam Boyu Şekli-Türkçe Uyarlamasının Geçerlik ve Güvenirliği. Çocuk ve Gençlik Ruh Sağlığı Dergisi 11: 109-16.

Hisli N (1989) Beck Depresyon Envanterinin üniversite öğrencileri için geçerliği, güvenirliği. Psikoloji Dergisi 7: 3-13.

Irmak TY, Sütçü ST, Aydın A et al (2007) Otizm Davranış Kontrol Listesinin (ABC) geçerlik ve güvenirliğinin incelenmesi. Çocuk ve Gençlik Ruh Sağlığı Dergisi 14: 13-23.

Incekaş S (2009) Çocukluk Otizmini Derecelendirme Ölçeği Geçerlik ve Güvenirlik Çalışması. YayınlanmamışTıpta Uzmanlık Tezi, Dokuz Eylül Üniversitesi Tıp Fakültesi.

Kaner S, Büyüköztürk Ş, Işeri E (2006) Yenilenmiş conners ana baba derecelendirme ölçeği kısa Türkçe formunun psikometrik özellikleri. Uluslararası Gelişimsel Nöropsikiyatri Toplantıları, İstanbul, p. 80-1.

Kaufman J, Birmaher B, Axelson D et al (2016). Schedule for Affective Disorders and Schizophrenia for School-Aged Children: Present and Lifetime Version (K-SADS-PL) DSM-5 November 2016 working draft. New Haven, Yale University, Child and Adolescent Research and Education.

Kaufman J, Birmaher B, Brent D et al (1997) Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. J Am Acad Child Adolesc Psychiatry 36: 980-8.

Kaufman J, Townsend LD, Kobak K (2017) The Computerized Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS): Development and Administration Guidelines. J Am Acad Child Adolesc Psychiatry 56: 357.

Kim YS, Cheon KA, Kim BN et al (2004) The reliability and validity of kiddie-schedule for affective disorders and schizophrenia-present and lifetime version- Korean version (K-SADS-PL-K). Yonsei Med J 45: 81-9.

Kocael Ö (2015) Çocuk ve ergenlerde irritabilite: Duygusal reaktivite indeksi'nin Türkçe geçerlilik güvenilirlik çalışması. Yayınlanmamış Tıpta uzmanlık tezi, Bursa, Uludağ Üniversitesi, Tıp Fakültesi.

Kolaitis G, Korpa T, Kolvin I et al (2003) Schedule for affective disorders and schizophrenia for school-age children-present episode (K-SADS-P): a pilot inter-rater reliability study for Greek children and adolescents. Eur Psychiatry 18: 374-5.

Krug DA, Arick JR, Almond PA (1980) Behavior checklist for identifying severely handicapped individuals with high levels of autistic behavior. J Child Psychol Psychiatry 21: 221-9.

Landis J, Koch G (1977) The measurement of observer agreement for categorical data. Biometrics 33: 159-74.

Leffler JM, Riebel J, Hughes HM (2015) A Review of Child and Adolescent Diagnostic Interviews for Clinical Practitioners. Assessment 22: 690-703.

Lempp T, de Lange D, Radeloff D et al (2012) The clinical examination of children, adolescents and their families. IACAPAP e-Textbook of Child and Adolescent Mental Health. Rey JM (Ed), Geneva, International Association for Child and Adolescent Psychiatry and Allied Professions, p. 1-25.

Machado JD, Caye A, Frick PJ, Rohde LA (2013) DSM-5. Major changes for child and adolescent disorders. IACAPAP e-Textbook of Child and Adolescent Mental Health. Rey JM (Ed), Geneva, International Association for Child and Adolescent Psychiatry and Allied Professions, p. 1-22.

McHugh ML (2012) Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 22: 276–82.

de la Peña FR, Rosetti MF, Rodríguez-Delgado A et al(2018a) Construct validity and parent-child agreement of the six new or modified disorders included in the Spanish version of the Kiddie Schedule for Affective Disorders and Schizophrenia present and Lifetime Version DSM-5 (K-SADS-PL-5) J Psychiatr Res 101: 28-33.

de la Peña FR, Villavicencio LR, Palacio JD et al (2018b) Validity and reliability of the kiddie schedule for affective disorders and schizophrenia present and lifetime version DSM-5 (K-SADS-PL-5) Spanish version. BMC Psychiatry 14: 193.

Regier DA, Narrow WE, Clarke DE et al (2013) DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses. Am J Psychiatr 170: 59–70.

Rettew DC, Lynch AD, Achenbach TM et al (2009) Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. Int J Methods Psychiatr Res 18: 169-84.

Shahrivar Z, Kousha M, Moallemi S et al (2010) The reliability and validity of kiddie-schedule for affective disorders and schizophrenia-present and life-time version-Persian version. Child Adolesc Ment Health 15: 97-102.

Shanee N, Apter A, Weizman A (1997) Psychometric properties of the K-SADS-PL in an Israeli adolescent clinical population. Isr J Psychiatry Relat Sci 34: 179-86.

Schopler E, Reichler RJ, DeVellis RF et al (1980) Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). J Autism Dev Dis 10: 91-103.

Stringaris A, Goodman R, Ferdinando S et al (2012) The Affective Reactivity Index: A concise irritability scale for clinical and research settings. J Child Psychol Psychiatry 53: 1109–17.

Sucuoğlu B, Öktem F, Gökler B (1996) Otistik çocukların değerlendirilmesinde kullanılan ölçeklere ilişkin bir çalışma. Psikiyatri, Psikoloji, Psikofarmakoloji 4: 117-21.

Turkish Statistical Institute (2015) Gelir ve Yaşam Koşulları Araştırması Bölgesel Sonuçları. Türkiye Istatistik Kurumu Haber Bülteni 21824: 1-2.