

Reliability and validity of the self-report version of the Early Adolescent Temperament Questionnaire – Revised (EATQ-R) Short Form in a Turkish sample

Dilara Demirpence ¹ and Samuel Putnam²

¹Department of Child and Adolescent Mental Health, Şişli Hamidiye Etfal Research and Training Hospital, Turkey, ²Department of Psychology, Bowdoin College, Brunswick, Maine, USA

Abstract: This paper presents the first study to examine the psychometric qualities of a Turkish translation and adaptation of the Early Adolescent Temperament Questionnaire – Revised (EATQ-R). Participants included 1,109 Turkish adolescents (48.7% females) aged 11 to 15 years. Results showed that internal consistency of the broad EATQ-R factors ranged from .65 to .74, internal consistency of the 11 fine-grained scales was comparable to that reported for the original English instrument, and test–retest stability of EATQ-R scales ranged from .60 to .70. Principal axis factoring of the scales revealed a three-factor structure (Effortful Control, Surgency, and Affiliativeness) that roughly reflected the structure obtained in U.S. samples. Item-level factor analysis revealed and confirmed eight factors consistent with a priori scales. Concurrent validity, assessed through correlations between EATQ-R and the Strengths and Difficulties Questionnaire (SDQ) subscales, found that Effortful Control was negatively correlated with Total Difficulties and Hyperactivity-Inattention Problems; Effortful Control, Affiliativeness, and Surgency were positively correlated with Prosocial Behaviors; and Frustration was linked to Emotional, Conduct, Hyperactivity-Inattention, and Total Difficulties. Our results suggest that the EATQ-R is a useful questionnaire to measure reactive and regulative temperamental traits in Turkish children and adolescents.

Keywords: children and adolescents; early adolescent questionnaire; temperament

Correspondence to: Dr. Dilara Demirpence, Department of Child and Adolescent Mental Health, Sisli Hamidiye Etfal Research and Training Hospital, Şişli, 4371, Istanbul, Turkey. Email: dilarademirpence@gmail.com

Received 26 August 2018. Accepted 9 September 2019.

In a recent review, temperament scholars who adopted different approaches arrived at a common definition of *temperament* as consisting of early emerging basic dispositions in the domains of activity, affectivity, attention, and self-regulation; and suggested that these dispositions are the product of complex interactions among genetic, biological, and environmental factors (Shiner et al., 2012). This characterization is consistent with the psychobiological framework associated with Rothbart, in which temperament is described as individual differences in constitutionally based reactivity and self-regulation dimensions (Putnam, Ellis, & Rothbart, 2001). *Reactivity* refers to excitability of the neuroendocrine, autonomic, and affective systems; and to individual differences in the thresholds of responsiveness,

reaction intensity, and responding time to reaction. While these reactions are regulated involuntarily in infancy, over time individuals increasingly gain the ability to regulate their reactions voluntarily, inhibiting and activating behaviors and controlling attention in response to the demands of a situation.

The Early Adolescent Temperament Questionnaire – Revised (EATQ-R) was developed by Ellis and Rothbart (2001) to measure reactive and regulatory temperamental traits in youths aged 9 to 15 years. The Short Form of this questionnaire includes 10 discrete subscales measuring temperament dimensions. The reliability and validity of the questionnaire has been studied in various languages (Hoffmann, Pérez, García, Rojas, & Martínez, 2017; Hsu,

2011; Muris & Meesters, 2009; Viñas Poch, 2015). In the original sample with which the EATQ-R was developed, the internal consistency reliability of the subscales ranged between .82 and .67 (Ellis & Rothbart, 2001). In a second U.S. sample, alpha coefficients of the subscales were reported between .43 and .82, whereas alphas for the factor-level scales ranged from .70 to .81 (Ellis, 2002). In a Dutch sample, after removing items with unsatisfactory total-item correlations, the internal consistency reliability of the EATQ-R subscales ranged between .61 and .73 (Muris & Meesters, 2009). In a Taiwanese study, internal consistency coefficients of the EATQ-R Short Form scales were found to be between .35 and .75 (Hsu, 2011). Finally, in a large sample ($n = 687$) of Chinese youth, alpha coefficients were between .47 and .80. Only two studies have examined test-retest reliability, finding correlations between .55 and .85 for the subscales in a Dutch sample and .62 and .72 in a Shanghai Chinese sample (Muris & Meesters, 2009; Zhang, Shen, & Gao, 2008).

The factor structure of the EATQ-R is somewhat ambiguous. The initial report by Ellis and Rothbart (2001) suggested four factors: Surgency (high-intensity pleasure, lack of fear, lack of shyness), Effortful Control (attention control, inhibitory control, and activation control), Affiliativeness (affiliation, perceptual sensitivity, and pleasure sensitivity), and Negative Affectivity (frustration). Muris and Meesters (2009) extracted three factors, obtaining a similar structure, with frustration loading on the Effortful Control factor, and Fear loading highly on all three dimensions. Item-level factor analysis has indicated a structure bearing only rough resemblance to the factor-level solutions. For instance, in a sample of Taiwanese adolescents, Hsu revealed four factors: an Effortful Control factor contained primarily items from attention control, inhibitory control, and activation control; but in contrast to Ellis and Rothbart (2001) and Muris and Meesters (2009), fear and frustration items formed a Negative Affect Factor, low-intensity pleasure items comprised their own factor, and affiliation items combined with high-intensity pleasure items to form a fourth. In the current study, factor analyses were carried out at both the scale- and item-level. The scale-level analyses allow more direct comparisons to the majority of the literature on the factor structure of the EATQ-R and related instruments (e.g., Putnam et al., 2001; Rothbart, Ahadi, Hershey & Fisher, 2001). Item-level analyses are useful for revealing information regarding specific items that are not clearly affiliated with a single dimension,

and present an alternate mode for exploring the underlying structure of the questionnaire.

Temperamental traits are related to psychosocial adjustment in childhood and adolescence. Using instruments created in the psychobiological model, Gartstein, Putnam, and Rothbart (2012) found high Negative Emotionality and low Effortful Control predicted both internalizing and externalizing problems in toddlers and preschool-aged children, with high Surgency linked to higher externalizing but lower internalizing. Similar findings were obtained using the EATQ-R by Oldehinkel, Hartman, DeWinter, Veenstra, and Ormel (2004). In that study, high shyness and low high-intensity pleasure, indicative of low Surgency, were characteristic of children high in Internalizing, but not Externalizing; frustration was linked to both Internalizing and Externalizing; and Effortful Control was low in children demonstrating externalizing problems, both with and without comorbid internalizing.

To our knowledge, only one study has explored links between temperament and adjustment in a Turkish population. Yilmaz et al. (2015) developed the Nine Type Temperament Scale, and used it to compare the temperamental traits of healthy adolescents and those with attention deficit hyperactivity disorder (ADHD). A cluster of children defined as impulsive, talkative, and extraverted was shown to be significantly higher in adolescents with ADHD. These findings complement those obtained by Muris and Meesters (2009) showing links between low EATQ-R effortful control and attention-hyperactivity problems in Dutch children.

The EATQ-R has been studied in Western European countries and Far Eastern countries; however, it has not been investigated in the region surrounding Turkey. Extending research to areas including the Middle East is important, as recent analyses of the literature indicate that 96% of psychological samples come from countries containing only 12% of the world's population, and predominantly from Western, educated, industrialized, rich, and democratic (WEIRD) societies (Henrich, Heine, & Noranyazan, 2010). Very limited research has been conducted concerning temperament profiles in a non-WEIRD country in this age range (Yilmaz et al., 2015), and this research has used instruments not commonly used elsewhere. Validation of relatively brief instruments may be particularly important, as easing demands on research participants may facilitate recruitment of large, representative samples in these understudied societies. With this in mind, in the current study the self-report version of the EATQ-R Short Form

was administered to a large sample of children and adolescents to study the following issues: (1) reliability (i.e., internal consistency reliability and test–retest stability) of the temperament scales; (2) replication of previous explorations of factor structure; and (3) construct validity of the EATQ-R temperament scales as revealed through associations with measures of psychopathology.

Methods

Participants and procedure

Young adolescents ($n = 1,109$; 48.7% female; $M_{\text{age}} = 12.61 \pm 1.12$ years) were recruited from three secondary schools in Istanbul, Turkey. The surveys were administered by school counselors in their classrooms. Prior to distributing the measures, the counselors explained the aims of the study and gave the adolescents information regarding informed consent, emphasizing that these questionnaires were not a test or exam and that they were not required to participate.

The majority of students enrolled in these schools (50%) participated in the initial data collection. Most adolescents attending these schools are native Turks, with less than 6% having emigrated from countries, such as Syria (World Migration Report, 2018). These young adolescents were not asked to provide socioeconomic information, but estimates provided by the school district suggest that these children come from families with average incomes for Istanbul (i.e., approximately \$5,900 U.S. per annum). Most parents in these families would have graduated from high school. A sizable number would have completed secondary school students aged 10–14 only (Turkish Statistical System and Turkstat Istanbul Regional Office, 2015).

This study was approved by the ethical committee of the Cerrahpaşa Faculty of Medicine. Permission for collecting data from young adolescents in schools was granted by the İstanbul Provincial Directorate of National Educational. The young adolescents were informed about the study and informed consent was obtained. Approximately 3 weeks after the initial collection, 139 adolescents agreed to complete the EATQ-R a second time.

EATQ-R

The self-report version of the EATQ-R Short Form (Ellis & Rothbart, 2001) was used. This instrument consists of 65 items regarding adolescents' feelings and behaviors

in response to commonly occurring situations, asking them to respond on a 5-point Likert-type scale ranging from *almost always true* to *almost always untrue*. These items form 12 scales. Two of these scales measure tendencies (Aggression and Depressive Mood) that are not conceptualized as temperament, and which are not analyzed in the current report. The 10 temperament scales, their definitions, the number of items, and example items, are as follows: (1) Activation Control (the capacity to perform an action when there is a strong tendency to avoid it; five items; "I finish my homework before the due date"); (2) Affiliation (the desire for warmth and closeness with others, independent of shyness or extraversion; five items; "I want to be able to share my private thoughts with someone else"); (3) Attention (the capacity to focus attention as well as to shift attention when desired; six items; "It is easy for me to really concentrate on homework problems"); (4) Fear (unpleasant affect related to anticipation of distress; six items; "I worry about getting into trouble"); (5) Frustration (negative affect related to interruption of ongoing tasks or goal blocking; seven items; "I get very upset if I want to do something and my parents won't let me"); (6) High-Intensity Pleasure (pleasure derived from activities involving high intensity or novelty; six items; "I find the idea of driving a race car exciting"); (7) Inhibitory Control (the capacity to plan, and to suppress inappropriate responses; five items; "When someone tells me to stop doing something, it is easy for me to stop"); (8) Perceptual Sensitivity (detection or perceptual awareness of slight, low-intensity stimulation in the environment; four items; "I notice even little changes taking place around me, like lights getting brighter in a room"); (9) Low-Intensity Pleasure (amount of pleasure related to activities or stimuli involving low intensity, rate, complexity, novelty, and incongruity; five items; "I like to look at trees and walk amongst them"); and (10) Shyness (behavioral inhibition to novelty and challenge, especially social; four items; "I feel shy about meeting new people"). As described in the Introduction, these scales can be organized into three factors: Surgency (high-intensity pleasure, lack of fear, lack of shyness), Effortful Control (attention control, inhibitory control, and activation control), and Affiliativeness (affiliation, perceptual sensitivity, and low-intensity pleasure). Scale scores are calculated as the mean score of scale items and factor scores are calculated as the mean of applicable scale scores.

The translation of the form into Turkish was carried out in three stages. In the first stage, the EATQ-R was

translated by two child psychiatrists independently. In the second stage, these translations were compared by the two child psychiatrists, who then chose the translation deemed most clear. This Turkish form was translated into English by a child psychologist with strong English skills. The back translation was examined by the administrator of the Rothbart Measures website, who identified and corrected inconsistencies between the back translation and the original English form. Corrections to the Turkish form were made on the basis of this feedback. Before collecting data, the questionnaire was reviewed by 10 adolescents assisting at the Child and Adolescent Mental Health Department of the Cerrahpaşa Faculty of Medicine. No items were identified as difficult to understand and no additional changes were made at this stage. This translation was submitted to the Istanbul Provincial Directorate of National Education for approval. The Provincial Directorate determined that three items were not acceptable for early Turkish adolescents. In response to these concerns, these items were modified. These changes are shown in Table 1.

Strengths and Difficulties Questionnaire

The Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001) consists of 25 items describing positive and negative attributes of children and adolescents that are allocated to five subscales of five items each: Emotional Symptoms (e.g., “unhappy”), Conduct Problems (e.g., “fights”), Hyperactivity-Inattention (e.g., “fidgety”), Peer Problems (e.g., “bullied”), and Prosocial Behavior (e.g., “considerate”). Each item is scored on a three-point scale: 0 = *not true*,

Table 1

EATQ-R Items That Were Slightly Changed by Istanbul Directorate of National Education

Original (English) item	Initial back translation	Adapted item back translation
When I am angry, I throw or break things.	When I am angry, I throw or break things.	When I am angry I harm things.
If I get really mad at someone, I might hit them.	If I really get angry with someone, I might hit her/him.	If I really get angry with someone, I might harm her/him.
I am nervous of some of the kids at school who push people into lockers and throw your books around.	I get nervous with people who push others and mess up with their books at school.	I get nervous with people who harm others and mess up with their books at school.

1 = *somewhat true*, and 2 = *certainly true*. Subscale scores are computed by summing scores on relevant items (after recoding reversed items; range 0–10) and a Total

Difficulties subscale is created by summing four subscales (Emotional Symptoms, Conduct Problems, Hyperactivity-Inattention, and Peer Problems). In the current study, we used the Turkish version, which has been shown to be reliable and valid in other Turkish samples (Yalın, Özbek, Güvenir, & Baydur, 2013). In the current study, Cronbach's alphas for the five subscales and the Total Difficulties scale were as follows: Emotional Symptoms (.62), Conduct Problems (.55), Hyperactivity-Inattention (.49), Peer Problems (.23), Prosocial Behavior (.66), and Total Difficulties (.61).

Statistical Analysis

SPSS was used for statistical analysis. As a first step, in order to compare the reliability of the Turkish EATQ-R scales with the original EATQ-R, Cronbach's alpha values were calculated for the 11 subscales and three factors. Test–retest stability was assessed with Pearson correlations. Factor structure was explored using principal axis factoring (PAF) of scale scores with an oblimin rotation, in order to provide comparability with the analyses conducted by Ellis and Rothbart (2001) and Muris and Meesters (2009). In order to investigate concurrent construct validity, we examined relationships between EATQ-R and SDQ scale scores using Pearson's correlations.

Results

Internal Consistency

As shown in Table 2, the internal consistency reliability of the EATQ-R scales ranged between .35 and .78, with five of the 12 scales demonstrating alphas lower than .60. These coefficients are only slightly lower than those reported by Ellis (2002), who obtained alphas between .43 and .82, including three scales with alphas below .60. When alpha was calculated at the factor scale level, by entering all items on the scales associated with Effortful Control, Affiliativeness, and Surgency, coefficients were .74, .76, and .66, respectively, which are also comparable to the values of .81, .81, and .70 reported by Ellis (2002).

Test–retest Stability

As shown in Table 2, stability of the subscales was between .47 and .79. The test–retest reliability of most subscales was similar to that obtained in the other known study of

Table 2*Internal Consistency Coefficients and Test–Retest Stability of A Priori EATQ-R Subscales*

EATQ-R Subscales	Cronbach's Alpha Turkish sample (<i>n</i> = 1109)	Test–retest stability		
		U.S. sample ¹	Turkish sample (<i>n</i> = 139)	Dutch sample ²
Activation Control	.66	.64	.47	.76
Attention Control	.48	.55	.57	.70
Inhibitory Control	.35	.43	.65	.78
Frustration	.64	.82	.68	.85
High-Intensity Pleasure	.42	.53	.67	.69
Shyness	.78	.75	.79	.73
Fear	.57	.61	.70	.73
Affiliation	.72	.65	.50	.80
Low-Intensity Pleasure	.70	.72	.69	.73
Perceptual Sensitivity	.62	.63	.59	.55

¹ Adapted from Ellis (2002).² Adapted from Muris and Meesters (2009).

EATQ-R stability, although the stability levels of the Activation Control, Frustration, and Affiliation subscales were considerably lower in the Turkish sample than in the Dutch sample collected by Muris and Meesters (2009). In the original EATQ-R studies, test–retest reliability was not investigated (Ellis, 2002; Ellis & Rothbart, 2001).

Factor Structure

Scale-level factor analysis

Following Ellis et al. (2001), exploratory factor analyses of scales were carried out using PAF with oblimin rotation. Examination of the scree plot suggested three factors, which accounted for 55.29% of the total variance. As shown in Table 3, the Activation Control, Attention Control, and Inhibitory Control subscales loaded on the first factor, which can be defined as Effortful Control.

Frustration loaded on both the first and the second factor. This second factor, Affiliativeness, also demonstrated high loadings by affiliation, perceptual sensitivity, pleasure sensitivity, and fearfulness. Fearfulness also loaded highly (negatively) on the third factor (Surgency), which was further defined by high-intensity pleasure and shyness (loading negatively). This structure is largely consistent with that obtained by Muris and Meesters (2009).

Item-level Factor Analysis

Following procedures used by Hsu (2002), the item scores were then subject to the PAF method with oblimin rotation, and subsequently refined to achieve simple structure by elimination of items with loadings under .3 for all factors, or over .3 for multiple factors. Eigenvalues for the first 12 factors were 4.57, 3.00, 2.74, 1.85, 1.64, 1.3, 1.20, 1.12,

Table 3*Scale-Level Factor Structure of Turkish EATQ-R Short Form*

EATQ-R subscales	Factors					
	A priori scale analyses			Derived scale analyses		
	Effortful Control	Affiliativeness	Surgency	Factor 1	Factor 2	Factor 3
Attention Control ¹	.65					
Activation Control ²	.61			.48	-.33	
Inhibitory Control ¹	.52					
Frustration	-.50	.47			.79	
Affiliation		.60		.60		
Perceptual Sensitivity ³		.55		.55		
Low-Intensity Pleasure		.45		.56		
High-Intensity Pleasure			.55			.56
Fearfulness		.46	-.52	.36		-.41
Shyness			-.35			

For reasons of clarity, only factor loadings > .35 are shown.

¹ Attention Control and Inhibitory Control not included in derived scale analyses.

² Derived Activation Control scale also contains one Inhibitory Control and one Attention Control item.

³ Derived Perceptual Control scale also contains one Inhibitory Control item.

0.94, 0.92, 0.90, and 0.87. Because the last substantial decrease was observed following the eighth factor, this solution was examined further.

The loadings for the eight-factor solution, containing 35 items that explain 49.75% of the variance, are shown in Table 4. With three exceptions, each factor contained items from a single scale. The eight factors, in order, primarily contain items from the Activation Control, Shyness, Frustration, High-Intensity Pleasure, Pleasure Sensitivity, Perceptual Sensitivity, Affiliation, and Fear subscales. The first factor (Activation Control) additionally included one item each from the Attentional Control and Inhibitory Control subscales; and the sixth factor (Perceptual Sensitivity) also contained an item from Attentional Control. Alternate forms of this model were then subject to a confirmatory factor analysis. Consistent with

Hsu (2002), one model considered all factors as orthogonal, while a second allowed the eight factors to correlate. Fit was evaluated with multiple indices, including chi-square, root-mean-square error of approximation (RMSEA), goodness-of-fit index (GFI), and comparative fit index (CFI). Chi-square typically rejects models when large sample sizes are used. Commonly used guidelines hold that CFI and GFI greater than .90, and RMSEA less than .05 indicate good fit (Baumgartner & Hombur, 1996; Hu & Bentler, 1999). The fits of these models are indicated in Table 5. The orthogonal model demonstrated less-than-good fit: $\chi^2(560) = 2280$, RMSEA = .053, GFI = .878, CFI = .744. The correlated model surpassed the thresholds for RMSEA and GFI, and neared the threshold for CFI: $\chi^2(532) = 1337.74$, RMSEA = .037, GFI = .932, CFI = .894.

Table 4
Item-Level Factor Structure of Turkish EATQ-R

Item number	Subscale	1	2	3	4	5	6	7	8
49	Activation Control	.68							
39	Activation Control	.66							
1	Attention Control	.55							
30	Activation Control	.51							
63	Inhibitory Control	.43							
18	Activation Control	.37							
7	Activation Control	.36							
45	Shyness		.92						
53	Shyness		.83						
15	Shyness		.67						
8	Shyness		.35						
47	Frustration			.66					
36	Frustration			.51					
25	Frustration			.49					
56	Frustration			.48					
62	Frustration			.44					
48	High-Int. Pleasure				.73				
42	High-Int. Pleasure				.56				
28	High-Int. Pleasure				.47				
65	Low-Int. Pleasure					.70			
33	Low-Int. Pleasure					.65			
16	Low-Int. Pleasure					.63			
23	Low-Int. Pleasure					.49			
6	Perceptual Sensitivity						.64		
12	Perceptual Sensitivity						.60		
21	Perceptual Sensitivity						.46		
41	Attention Control						.44		
44	Affiliation							.52	
17	Affiliation							.45	
27	Affiliation							.44	
54	Affiliation							.38	
31	Affiliation							.33	
35	Fear								.49
51	Fear								.46
40	Fear								.36

For reasons of clarity, only factor loadings > .35 are shown.

Table 5
Correlations Between EATQ-R Factors and SDQ Subscales

	SDQ Subscales					
	Emotional	Conduct	Hyperactive	Peer	Prosocial	Total
Factors						
Effortful Control	-.31**	-.43**	-.56**	-.17**	.42**	-.53**
Affiliativeness	.03	-.21**	-.18**	-.06*	.48**	-.17**
Surgency	-.38**	.09**	-.13**	-.13**	-.10**	-.19**
Frustration	.25**	.23**	.31**	-.02	-.01	.29**
Derived scales						
Activation Control ¹	-.20**	-.34**	-.45**	-.12**	.40**	-.40**
Shyness	.34**	.03	.15**	.10**	.04	.23**
Frustration	.22**	.30**	.37**	.00	-.12**	.32**
High-Int. Pleasure	-.13**	.17**	-.01	.02	-.02	.01
Low-Int. pleasure	.04	-.20**	-.18**	-.04	.37**	-.13**
Perceptual sensitivity ²	.02	-.09**	-.13**	-.07*	.29**	-.09**
Affiliation	.01	-.16**	-.10**	-.18**	.41**	-.15**
Fear	.21**	-.12**	.07*	-.03	.23**	.06

¹ Activation Control scale also contains one Inhibitory Control item and one Attentional Control item.

² Perceptual Sensitivity scale also contains one Attentional Control item.

* $p < .05$. ** $p < .01$.

Scales were then created by averaging the item scores associated with each factor. Cronbach's alphas for these scales equaled .73, .78, .67, .61, .73, .64, .59, and .51. These eight scales were subject to exploratory PAF. Eigenvalues were 2.00, 1.35, 1.26, 0.90, 0.75, 0.64, 0.58, and 0.53. Three factors were rotated. As shown in Table 3, Factor 1 demonstrated loadings over .35 for five scales: Activation Control, Low-Intensity Pleasure, Perceptual Sensitivity, Affiliation, and Fear. The second had a positive loading for Frustration and a negative loading for Activation Control. The third had a positive loading for High-Intensity Pleasure and a negative loading for Fear. The scale for Shyness did not load over .35 on any factor.

Concurrent validity of factor scores

For the sake of simplicity, to provide comparability with prior studies, and to allow use of robust and internally consistent scales, we did not use a priori subscale scores to examine relations between temperament and behavior problems, but instead used factor scores calculated as the average of scale scores affiliated with the factors. Because frustration formed its own "Negative Affectivity" factor in Ellis and Rothbart (2001), and loaded on two factors in the current study, it was left as its own score in relation to the SDQ subscales. However, because fearfulness was considered part of Surgency by Ellis and Rothbart (2002), it was used in the creation of the Surgency score used in the current study.

Correlations between EATQ-R factors and SDQ subscales are displayed in Table 5. Most convincingly, correlations were found between Effortful Control and

Hyperactivity-Inattention and Total Problems. Conduct Problems was negatively linked to Effortful Control and Affiliativeness. SDQ Prosocial Behavior was positively associated with Effortful Control and Affiliativeness. As expected, Affiliativeness was also significantly correlated with Prosocial Behavior. A negative association was found between Surgency and Emotional Problems. Furthermore, the Frustration scale was positively associated with Emotional, Conduct, and Hyperactivity-Inattention Problems. The results of the analysis demonstrated that all of the temperamental factors and the Frustration scale were correlated with the SDQ Total Difficulties.

Concurrent validity of scales derived from item-level factoring

Table 5 also contains correlations between SDQ scores and the scales developed on the basis of our item-level factor analyses. Because the factor structure of these scales was difficult to interpret (e.g., five of the eight scales loaded on a single factor; the other two factors were defined by only two scales), the factors from the exploratory exercise involving these scales were not analyzed in relation to SDQ scores. Patterns of correlations were reminiscent of those found with the factors derived from the a priori scales. That is, the correlations from Low-Intensity Pleasure, Perceptual Sensitivity, and Affiliation to SDQ scores were highly similar to those between the Affiliation factor and SDQ scores; and correlations between Activation Control and SDQ scores corresponded closely to those between the Effortful Control

factor and SDQ scores. Some distinctions were apparent with respect to the dimensions associated with Surgency. We describe these briefly in the following discussion.

Discussion

This is the first study to investigate the psychometric properties of the EATQ-R Short Form in a community sample of Turkish children and adolescents. Our initial analyses concerned the internal consistency reliability of both the fine-grained scales of the EATQ-R and of the items making up the broad factors. The internal consistency coefficients for the broad factors were reasonably high, suggesting their usefulness for studying temperament in Turkish adolescents. Less confidence is warranted with respect to the subscales, as Cronbach's alpha for five of these scales was under .60. Although .70 is widely considered a cut-off point for acceptable internal consistency (George & Mallery, 2003; Nunnally, 1978), this cut-off point has also been criticized as arbitrary (Goodwin & Goodwin, 1999), with some scholars ranking alphas of .60 as undesirable, but not acceptable, especially for short scales (De Vellis, 1991). Furthermore, these values are within norms for the topic area. In a recent review of temperament questionnaire methodology, of 43 instruments summarized, 28 contained at least one scale with an internal consistency estimate lower than .70, and 15 included at least one scale with a reported internal consistency less than .60 (Gartstein, Bridgett, & Low, 2012). In addition, in Taiwanese adolescents, Hsu (2011) found that internal consistency of the original EATQ-R scales was between .35 and .77, and Chang (2004) reported internal consistency of the EATQ-R Long Form scales to range between .47 and .69 (Chang, 2004). As such, the values obtained in our Turkish sample are reminiscent of those obtained with other translations.

To some degree, low alphas for the Short Form scales are to be expected in any sample, as internal consistency of a scale is inherently inversely related to the number of items in the scale (Gliem & Gliem, 2003), such that four-item scales would have to be considerably redundant to achieve adequate internal consistency. It is also of note that the three scales (Inhibition Control, Attention Control, and High-Intensity Pleasure) demonstrating the lowest internal consistency include reverse-scored items, which may have proven difficult to understand in the Turkish translation. Due to the questionable internal consistency of these scales,

we advise that researchers with specific interests in these fine-grained temperament attributes consider using the scales from the EATQ-R Long Form, rather than the Short Form, and that researchers whose questions do not necessitate specific traits use scores calculated at the factor level.

The current study confirmed that the Turkish EATQ-R Short Form had adequate stability. In the original EATQ-R studies, test-retest reliability was not investigated (Ellis, 2002). In a Shanghai Chinese sample, test-retest stability of the EATQ-R Short Form scales ranged between .62 and .72 (Zhang et al., 2008).

Our PAF of the 10 scales of the EATQ-R Short Form accounted for over 50% of the total variance, which is higher than the 41% that has been suggested as acceptable (Kline, 2014). More importantly, this analysis yielded three factors that were consistent with those obtained in other investigations of the structure of the EATQ-R (e.g., Muris & Meesters, 2009). Our item-level PAF yielded a structure that indicated factors associated with eight of the 10 a priori EATQ-R scales. As expected, given the low internal consistency of the Attention Control and Inhibitory Control scales, factors representing these dimensions were not revealed in the item-level analysis. When scales associated with these eight factors were themselves subject to a scale-level factor analysis, Activation Control (Activation Control) continued to load on the same factor as Frustration; the Surgency scales of High-Intensity Pleasure and Fear continued to load together; and the three Affiliativeness scales loaded together. As such, this exploratory analysis revealed a factor structure bearing some similarity to the three-factor model that has emerged from factor analyses of other fine-grained temperament measures (e.g., Putnam et al., 2001). Removal of two of the three Effortful Control scales did, however, have implications for the factor structure, as Activation Control was shown to load along with the Affiliativeness scales, a pattern not apparent when Inhibitory and Attentional Control were included in the analyses.

The broad factors of the Turkish EATQ-R demonstrated convergent validity through an expected pattern of relations with scales on a widely used measure of adolescent behavior problems. Similar to previous findings (e.g., Gartstein, 2012; Muris & Meesters, 2009; Oldehinkel et al., 2004), low Effortful Control was implicated in multiple types of problematic behavior, with particularly high relations to hyperactivity, and was also strongly linked to poor peer relations. Frustration also contributed to conduct problems,

hyperactivity problems, emotional problems, and total difficulties. Surgency appeared to be inversely related to emotional problems, a finding somewhat similar to that obtained by Gartstein (2012), who found that low Surgency was associated with low Internalizing. Curiously, in the current study, Surgency demonstrated only a low (positive) correlation with conduct problems and a low negative correlation with attention problems. These findings are inconsistent with prior studies linking Surgency to externalizing, and suggest that the implications of energetic traits may differ from culture to culture. Finally, Affiliativeness was modestly associated with low levels of behavior problems, but strongly predictive of prosocial behavior. More fine-grained analyses, carried out with the empirically derived scales resulting from our item-level factor analyses, revealed nuance in patterns of relations between temperament and behavior problems. For instance, shyness and fear were more strongly associated with emotional problems than high-intensity pleasure, whereas high-intensity pleasure was more closely linked to conduct problems. These differing implications for dimensions from a common factor indicate the value of considering narrow, empirically distinct dimensions in investigations of temperament in relation to other constructs.

In summary, we found that the EATQ-R is applicable to Turkish young adolescents. The broad factors and several subscales demonstrated adequate internal consistency, despite the brevity of the scales; test-retest reliability analyses confirmed the short-term stability of scales; and the factor structure revealed was consistent with past studies, as were patterns of relations with behavior problems.

Despite our satisfaction that the Turkish adaptation of the EATQ-R adequately achieved the psychometric properties of the original English form and other translations, this satisfaction is couched in more general concerns regarding the instrument. In multiple investigations (e.g., Ellis, 2002; Hsu, 2002), including ours, the internal consistency of several subscales was below acceptable levels. This was true both of the a priori, rationally derived scales and the very short (e.g., three items) scales created on the basis of our item-level factor analyses. Researchers have typically addressed this concern by using factor scores, rather than individual subscales, in their analyses. Although we recommend using this approach for the Turkish EATQ-R, it is not ideal: As indicated in our exploration of convergent validity, fine-grained scales can reveal important detail regarding implications of temperament for other

phenomena. In addition, the EATQ-R does not include several relevant dimensions that have been included in analogous measures for younger children (see Putnam et al., 2001). This has led to confusion regarding the factor structure: Although Ellis et al. (2002) reported a four-factor structure, only the Frustration scale was included in the Negative Affectivity factor. The decision of Ellis et al. (2002) to force a fourth factor was influenced by the regular inclusion of Negative Affectivity in the structure of analogous measures, and subsequent studies have often failed to replicate the four-factor structure. We feel that the Turkish adaptation of the EATQ-R is an acceptable tool for assessing temperament in Turkish adolescents, but encourage further exploration of adolescent temperament more broadly. Generation of a larger number of internally consistent items is warranted to allow reliable scales for all dimensions, and a wider variety of dimensions should be assessed. These steps should be complemented by ongoing exploration of factor structure at the item level, with the results of such analyses used to refine the scales of a revised instrument.

More immediately and specifically, future directions concerning measurement include investigations of the EATQ-R Long Form in Turkish samples, and possible enhancement of internally inconsistent scales (e.g., Inhibitory Control) to improve their suitability to Turkish culture. With respect to sampling, a limitation of the current study is that the data were collected from a nonclinical sample of young adolescents. Thus, future research involving clinically referred young adolescents could contribute to understanding the relation between temperament and psychiatric symptoms, and how these relations may be impacted by culture.

Disclosure of conflict of interest

The authors declare there are no conflicts of interest.

References

- Baumgartner, H., & Hombur, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing*, 13, 139–161. [https://doi.org/10.1016/0167-8116\(95\)00038-0](https://doi.org/10.1016/0167-8116(95)00038-0)
- Chang, C. T. (2004). *Temperament in Chinese children: A comparison of gender and self/parental ratings* (Unpublished doctoral dissertation). University of Hawaii, Honolulu, HI.
- De Vellis, R. F. (1991). Scale development: Theory and applications. *Journal of Educational Measurement*, 31(1), 79–82.

- Ellis, L. K. (2002). *Individual differences and adolescent psychosocial development* (Unpublished doctoral dissertation). University of Oregon, Eugene, OR.
- Ellis, L. K., & Rothbart, M. K. (2001). *Revision of the Early Adolescent Temperament Questionnaire*. Poster session presented at the biennial meeting of the society for research in child development (). Minneapolis, MN.
- Gartstein, M. A., Putnam, S. P., & Rothbart, M. K. (2012). Etiology of preschool behavior problems: Contributions of temperament attributes in early childhood. *Infant Mental Health Journal*, 33(2), 197–211. <https://doi.org/10.1002/imhj.21312>
- Gartstein, M. A., Bridgett, D. J., & Low, C. M. (2012). Asking questions about temperament. In Z. Marcel & L. S. Rebecca (Eds.), *Handbook of temperament* (). The Guilford Press, New York, NY.
- George, D., & Mallery, M. (2003) Using SPSS for Windows step by step: A simple guide and reference Allyn & Bacon, Boston, MA.
- Gliem, J. A., & Gliem, R. R. (2003, October). *Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales*. Paper presented at the Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education, Columbus, OH.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337–1345. <https://doi.org/10.1097/00004583-200111000-00015>
- Goodwin, L. D., & Goodwin, W. L. (1999). Measurement myths and misconceptions. *School Psychology Quarterly*, 14(4), 408–427. <https://doi.org/10.1037/h0089017>
- Henrich, J., Heine, S. J., & Noranyazan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33, 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Hoffmann, M., Pérez, J. C., García, C., Rojas, G., & Martínez, V. (2017). Chilean adaptation and validation of the Early Adolescent Temperament Questionnaire-Revised Version. *Frontiers in Psychology*, 8, 2131. <https://doi.org/10.3389/fpsyg.2017.02131>
- Hsu, K. Y. (2011). Psychometric properties of the revised Early Adolescent Temperament Questionnaire in Taiwanese adolescents. *International Journal of Educational and Psychological Assessment*, 7, 19–33.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Kline, P. (2014). *An easy guide to factor analysis*. Routledge, England. <https://doi.org/10.4324/9781315788135>
- Muris, P., & Meesters, C. (2009). Reactive and regulative temperament in youths: Psychometric evaluation of the Early Adolescent Temperament Questionnaire-Revised. *Journal of Psychopathology and Behavioral Assessment*, 31(1), 7–19. <https://doi.org/10.1007/s10862-008-9089-x>
- Nunnally, J. (1978). *Psychometric methods*. McGraw-Hill, New York, NY.
- Oldehinkel, A. J., Hartman, C. A., De Winter, A. F., Veenstra, R., & Ormel, J. (2004). Temperament profiles associated with internalizing and externalizing problems in preadolescence. *Development and Psychopathology*, 16(2), 421–440.
- Putnam, S. P., Ellis, L. K., & Rothbart, M. K. (2001). The structure of temperament from infancy through adolescence. In A. Elias & A. Angleitner (Eds.), *Advances in research on temperament* (pp. 165–182). Pabst Scientist Publisher, Lengerich, Germany.
- Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child development*, 72(5), 1394–1408.
- Shiner, R. L., Buss, K. A., McClowry, S. G., Putnam, S. P., Saudino, K. J., & Zentner, M. (2012). What is temperament now? Assessing progress in temperament research on the twenty-fifth anniversary of Goldsmith et al. *Child Development Perspectives*, 6(4), 436–444.
- Turkish Statistical System and Turkstat Istanbul Regional Office. (2015). Study Visit of the UNSD, ESCAP and OSCE to Turkey 17 Sep 2015, İstanbul. Retrieved from https://unstats.un.org/unsd/demographic/meetings/wshops/Turkey/2015/-docs/Presentations/Session15-Turkstat_overview_ENG.pdf
- Viñas Poch, F. G. (2015). Psychometric properties of the EATQ-R among a sample of Catalan-speaking Spanish adolescents. *Universitas Psychologica*, 14(2), 747–758. <https://doi.org/10.11144/Javeriana.upsy14-2.ppea>
- World Migration Report 2018: *International organization for migration*, Geneva, SUI. https://publications.iom.int/fr/system/files/pdf/wmr_2018_en_chapter7.pdf
- Yalın, Ş., Özbek, A., Güvenir, T., & Baydur, H. (2013). The advanced psychometric properties of Turkish Strengths and Difficulties Questionnaire (SDQ). *Turkish Journal of Child and Adolescent Mental Health*, 20(1), 23–32.
- Yılmaz, E. D., Ünal, Ö., Palancı, M., Kandemir, M., Örek, A., Akkin, G., ... Selçuk, Z. (2015). Dokuz Tıp Mizaç Ölçeği Ergen (DTMÖ-E) Formunun Geçerlik-Güvenirliliği ve Ergenlerde Dikkat Eksikliği Hiperaktivite Bozukluğu İle Mizaç Tipleri Arasındaki İlişki. *EMİTİM VE BİLİM*, 40(179), 361–381.
- Zhang, J., Shen, L., & Gao, L. (2008). The revision and application of the revision of the Early Adolescent Temperament Questionnaire. *Chinese Mental Health Journal*, 22(6), 439–443.