Reliability and Validity Study of Turkish Version of Pediatric Eosinophilic Esophagitis Symptom Scores® (Tr-PEESS v2.0) Led to Development of a New Pediatric Eosinophilic Esophagitis Scale: GaziESAS

Hacer Ilbilge Ertoy Karagol[1], Dilek Yapar[2], Odul Egritas Gurkan[3], Sinan Sari[3], Mustafa Necmi Ilhan[2], Buket Dalgic[3,], Arzu Bakirtas[1,*], Gazi University Pediatric Eosinophilic Gastrointestinal Diseases Working Group*

[1]Department of Pediatric Allergy, Gazi University School of Medicine, Ankara, Turkey
[2]Department of Public Health, Gazi University School of Medicine, Ankara, Turkey
[3]Department of Pediatric Gastroenterology, Gazi University School of Medicine, Ankara, Turkey

Cite this article as: Ilbilge Ertoy Karagol H, Yapar D, Egritas Gurkan O, *et al*. Reliability and validity study of "Turkish" version of pediatric eosinophilic esophagitis symptom scores® (Tr-PEESS v2.0) led to development of a new pediatric eosinophilic esophagitis scale: GaziESAS. *Turk J Gastroenterol.* 2021; 32(4): 365-373.

ABSTRACT

Background: *There has been no valid and reliable Turkish scale that measures symptoms in children with eosinophilic esophagitis (EoE). The aim of the study is to test the validity and reliability of the Turkish version of Pediatric Eosinophilic Esophagitis Symptom Scores® (Tr-PEESS v2.0).*

Methods: *Relevant forms of Tr-PEESS v2.0 were applied to 2-18 years old children with EoE and to their parents. KINDL QoL patient and parent questionnaires and the GaziESAS scale developed in this study were used to test the convergent validity of Tr-PEESS v2.0. Discriminant validity was evaluated among 3 EoE treatment groups: under treatment, off treatment due to remission, and uncompliant with treatment. Reliability was evaluated by internal consistency, test–retest reliability, and item analysis.*

Results: *Fifty-two children/teens (mean age 130.2 ± 60.3 months) and 84 parents were interviewed twice one week apart. The mean duration of EoE was 47.2 ± 35.6 months. Tr-PEESS v2.0 reports correlated with GaziESAS (range 0.361-0.855) and KINDL QoL questionnaires (range −0.316 to 0.413). Parent report of Tr-PEESS v2.0 discriminated children uncompliant with treatment from the ones off treatment and undertreatment. Cronbach's α values and intraclass correlation coefficients (ICC) values of Tr-PEESS v2.0 ranged from 0.614-0. 895 and 0.646-0.910, respectively.*

Conclusion: *Tr-PEESS v2.0 is a valid and reliable tool to use in Turkish children. GaziESAS is a new parent-proxy pediatric EoE scale with an additional adaptive behavior domain that passed scale developmental stages successfully for Turkish children with EoE.*

Keywords: *Children, eosinophilic esophagitis, reliability, scale, validity*

INTRODUCTION

Eosinophilic esophagitis (EoE) is a chronic disorder characterized by esophageal dysfunction and eosinophilic inflammation of the esophagus. The most bothersome symptoms of EoE are food impaction and dysphagia, which are more prominent in adolescents and adults. In children and infants EoE may present with gastroesophageal reflux like symptoms, abdominal pain, and feeding problems.[1] Patients with EoE usually develop adaptive behaviors to overcome these symptoms, such as imbibe fluids with meals, modify food (cutting into small pieces, pureeing), prolong meal times, avoid hard texture foods, chew excessively, turn away tablets/pills.[2] Besides treatment issues, symptoms and adaptive behaviors may affect the quality of life and may constitute the main components of patient-reported outcomes (PRO). Standardized and validated instruments that measure PRO are necessary for clinical use and research purposes. There are 2 instruments in children with EoE that measure PRO: Pediatric Eosinophilic Esophagitis Symptom Scores® (PEESS v2.0) and Pediatric Quality of Life Inventory™ (PedsQL™) EoE Module.

PEESS™ v2.0 was developed in 2009 by Franciosi et al.[3] It is a content validated scale that identifies histologic and molecular correlates of the clinical features of EoE.[4] There

*These authors have contributed equally to this work.

Corresponding author: ***Hacer Ilbilge Ertoy Karagol****, e-mail:* ***theilbilge@yahoo.com.au***
Received: ***August 7, 2020*** *Accepted:* ***September 21, 2020***

DOI: ***10.5152/tjg.2020.20674***

has been no developed valid and reliable Turkish scale that measures PRO in children with EoE, yet. In addition, there is no study to evaluate the validation and reliability of PEESS v2.0 in Turkish. Therefore, we aimed to test the validity and reliability of the Turkish version of PEESS v2.0 (Tr-PEESS v2.0).

MATERIALS AND METHODS

Study Population

Children between 2 and 18 years old who have been followed up at our center with a diagnosis of EoE since 2008 and their parents were included in the study. EoE diagnosis was made according to the guideline.[5] All the forms used in the study were filled in by face-to-face interview with the patients and their parents at our center. These patients have regular control visits at 3-month intervals at our center; however, for the ones who do not attend regularly or whose control visit was out of the study period were invited to our center by telephone call. The patients were seen twice at our center, one week apart. At the first visit, demographics were filled in, and Tr-PEESS v2.0, GaziESAS, and KINDL QoL forms were applied to patients and to their parents. In the second visit, only Tr-PEESS v2.0 was applied.

DEFINITIONS

Treatment Groups

Undertreatment: It refers to children with EoE who are still on an EoE treatment for at least a month.

Off treatment: It refers to children who are in remission of EoE without any EoE treatment for at least a year and children who are under evaluation for remission without any EoE treatment.

Uncompliant with treatment: It refers to children with EoE who quitted EoE treatment by themselves or by parental decision.

Remission: It refers to children with EoE who are in remission symptomatically, endoscopically, and histopathologically for at least a year without any EoE treatment.

Main Points

- *There has been no developed valid and reliable Turkish scale that measures patient-reported outcomes (PRO) in children with eosinophilic esophagitis (EoE).*
- *Tr-PEESS v2.0 is a valid and reliable tool to use as PRO in EoE studies.*
- *GaziESAS is a high-quality parent-proxy scale with an additional adaptive behavior domain.*

Ethics

A written informed consent form was obtained from all parents and children ≥12 years old. This study was approved by Gazi University Ethics Committee (IRB No: 2019-312).

Questionnaires

Tr-PEESS V2.0

The rights of use, distribution, and linguistic translation of PEESS v2.0 were given to Mapi Research Trust by its developers.[3,4,6] A scale request was created to test the validity and reliability of Tr-PEESS v2.0. The license agreement was signed, and Tr-PEESS v2.0 was provided by Mapi Research Trust (Figure 1).[6]

Tr-PEESS v2.0 has a children/teens report (8-18 years old) and parent-proxy report for children/teens (2-18 years old). Tr-PEESS v2.0 assesses the frequency and severity of EoE-related symptoms. Both children/teens and parent-proxy forms consist of 20 items, 11 of which are related to the frequency (items number 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20), 9 of which are related to the severity (items number 2, 4, 6, 8, 10, 12, 14, 16, 18) domains. Frequency and severity domains are scored with a 5-point Likert-type scale ranging from "never" to "almost always" and "not bad at all" to "very bad" and with face figures expressing different emotional states, respectively. The item scores for both domains range from 0 to 4. The range of total/domain scores is transformed into a scale from 0 to 100 (0 = 0; 1 = 25; 2 = 50; 3 = 75; 4 = 100). Higher scores indicate more frequent and severe symptoms.

GaziESAS

There is no similar EoE scale developed in Turkey to test convergent validity (CgV) of Tr-PEESS v2.0. Therefore, we developed a new scale that we called Gazi University Eosinophilic esophagitis Symptoms and Adaptive behavior Scale (GaziESAS) to evaluate the symptoms and adaptive behaviors to overcome these symptoms in children with EoE in the last month.

Stages of Item Pool of GaziESAS

Five experts of our research team (AB, HIEK, OEG, SS, BD) determined the item pool of GaziESAS independently on their own by clinical experience and searched in related studies. This study decided first to develop a parent-proxy scale to assess children between 2 and 18 years old.

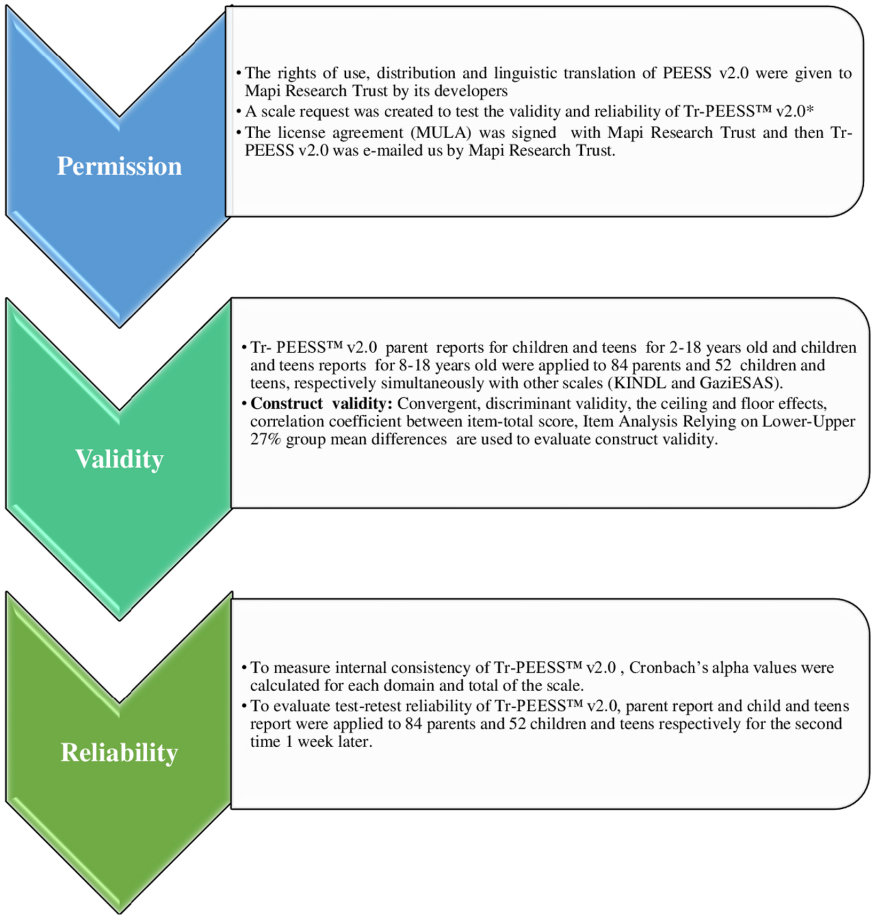

Figure 1. Flow chart of validity and reliability of Tr-PEESS v2.0. *Tr-PEESS v2.0: Turkish version of PEESS v2.0.

Item pool consisted of 57 items. The experts who created the item pool reviewed all the items, deleted duplications, and discussed their differences and similarities. After this synthesis, 52 items with 2 domains stayed in the pool (27 items for EoE symptoms and 25 items for adaptive behaviors domains), which were scored in 3-point Likert-type (0 = never, 1 = sometimes, and 2 = often). The steps of validity and reliability of GaziESAS are presented in the "Statistical analysis" section.

KINDL

We used Kinder Lebensqualität Fragebogen quality of life questionnaire (KINDL QoL) as another questionnaire to test convergent validity of Tr-PEESS v2.0. KINDL QoL is a generic scale and has patient and parent-proxy QoL Questionnaires.[7] Both questionnaires were tested for reliability and validity by Eser et al. in Turkey.[8] The questionnaires of KINDL QoL for children and adolescents consist of 24 items in the 6 subscales of "physical well-being, emotional well-being, self-esteem, family, and friends, school" and the last subscale consisted of 6 items for chronic disease, which measures the child's QoL with respect to his or her illness. Mean total/subscale scores are expressed between 0 and 100. High scores are indicators of high QoL.

Statistical Analysis

The SPSS program, version 22.0, was used to analyze the data. Normal distribution for continuous variables was assessed with visual (histograms and probability graphics) and analytic methods (Kolmogorov–Smirnov and Shapiro–Wilk tests). Statistical significance was accepted as $P < .05$.

Content Validity (CnV)

GaziESAS was sent to 8 experts in pediatric allergy and gastroenterology for assessment of CnV. The expert

committee was instructed to rate 52 items of Gazi-ESAS on a 4-point Likert scale. They scored on the comprehensibility, grammaticality, adequacy, clarity, and simplicity of GaziESAS items and made the necessary changes accordingly.[9]

Content validity ratio (CVR) is an item statistic based on CnV, and it helps the decision whether the item must stay in the scale or not.[10] The scale's content validity index (CVI) is calculated by taking the average of the CVR's of the items.[10] If CVI is > CVRcritical, the CnV of the scale is statistically significant.[10] Another expert opinion for significant CnV is to have a CVI > 0.8.[11,12]

Construct Validity (CsV)

Convergent and discriminant validity (DV) are both considered subcategories of CsV. Convergent validity (CgV) was found by Spearman correlation analysis of GaziESAS, KINDL, and Tr-PEESS v2.0 scores (correlation coefficient (CC): $r \geq 0.81$–1.0 excellent, 0.61-0.80 very good, 0.41-0.60 good, 0.21-0.40 fair, and 0-0.20 poor).[13] Fair to high correlation shows evidence of CgV.[14,15] In DV, the ability of the scale is tested to distinguish the groups of different structures and characters.[16,17] The Mann–Whitney *U* test was used to compare children's treatment groups with EoE for DV.

The floor and ceiling effects and CC between item-total scores are also recommended as scale validity and reliability indicators.[18,19] The CC between item-total score and item-subscale score should be at least 0.20[20] in a positive direction, and floor and ceiling effects should be under 0.15-20 to consider a scale homogeneous.[21]

Reliability: Internal Consistency

Cronbach's alpha values per domain and for total were computed. A Cronbach's alpha value between 0.60 and 0.80 is regarded as evidence of good reliability and >0.80 as excellent reliability.[22,13]

Test–Retest Reliability

Retests of parent (n = 84) and child and teens (n = 52) reports of Tr-PEESS v2.0 were applied one week apart. Test–retest reliability was determined by using the intraclass correlation coefficient (ICC). ICC between 0.60 and 0.80 is regarded as evidence of good reliability and >0.80 as excellent reliability.[13]

RESULTS

Participants

Eighty-four children (25 girls, 59 boys) and their mothers were included in this study. Fifty-two children (61.9 %) were between 8 and 18 years old. The mean age of children was 130.2 ± 60.3 months (range, 32-216 months). The mean duration of EoE was 47.2 ± 35.6 months (range, 3-126 months). Thirty percent (n = 25) of our patients were from other cities. As allergic comorbidity, 15 children had asthma (17.9%), 22 had allergic rhinitis (26.2%), 9 had atopic dermatitis (10.7%), and 2 had food allergy (2.4%). Forty-seven children had a family history of allergic disease (56%). Thirty patients (35.7%) were under treatment (PPI: 14, swallowed topical corticosteroids: 4, elimination diet: 6, combined treatment: 6). Fifty-four patients (64.3%) were off EoE treatment (our team stopped 16 patients' treatment due to good disease control, 5 patients were in remission without any EoE treatments for at least a year, 33 patients stopped their treatment themselves "uncompliant" with treatment.

Psychometric Evaluation of GaziESAS

It is given as supplement files (S1, S2).

Psychometric Evaluation of Tr-PEESS v2.0

Construct Validity of Tr-PEESS v2.0

The correlations between domains and total scores and between parent and children/teens reports of Tr-PEESS v2.0 were good (Table 1). For CgV analysis, the CCs between Tr-PEESS v2.0, GaziESAS, and KINDL scales were calculated (Table 1). The correlations between Tr-PEESS v2.0 and GaziESAS scores were moderate to strong, and between Tr-PEESS v2.0 and KINDL, total scores were moderate. Not children/teens report, but parent report of Tr-PEESS v2.0 correlated with KINDL disease dimension scores for parent (Table 1).

The duration of EoE showed a negative moderate correlation with all parent Tr-PEESS v2.0 scores but only with frequency domain of child/teens report (Table 1).

Discriminant Validity of Tr-PEESS v2.0 and GaziESAS

The DV analyses of the Tr-PEESS v2.0 and GaziESAS were made by comparison of scores between treatment groups (Table 2). Tr-PEESS v2.0 parent report and GaziESAS scores were significantly higher in the uncompliant group compared to the off-treatment group ($p<0.05$), (Table 2). Scores of Tr-PEESS v2.0 parent but not child/teens report and that of GaziESAS were found significantly different between treatment groups (Table 2).

Table 1. Evaluation of Construct Validity of Tr-PEESS v2.0

	Tr-PEESS v2.0					
	CHILDREN AND TEENS REPORT			PARENT REPORT		
	Frequency Domain r (*P*)	Severity Domain r (*P*)	Total r (*P*)	Frequency Domain r (*P*)	Severity Domain r (*P*)	Total r (*P*)
Tr-PEESS v2 Children and Teens Report						
Frequency Domain	-	0.690 (<0.001)	0.890 (<0.001)	0.529 (<0.001)	0.688 (<0.001)	0.662 (<0.001)
Severity Domain	0.690 (<0.001)	-	0.932 (<0.001)	0.560 (<0.001)	0.499 (<0.001)	0.579 (<0.001)
Total	0.890 (<0.001)	0.932 (<0.001)	-	0.598 (<0.001)	0.636 (<0.001)	0.673 (<0.001)
Tr-PEESS v2 Parent Report						
Frequency Domain	0.529 (<0.001)	0.560 (<0.001)	0.598 (<0.001)	-	0.754 (<0.001)	0.899 (<0.001)
Severity Domain	0.688 (<0.001)	0.499 (<0.001)	0.636 (<0.001)	0.754 (<0.001)	-	0.955 (<0.001)
Total	0.662 (<0.001)	0.579 (<0.001)	0.673 (<0.001)	0.899 (<0.001)	0.955 (<0.001)	-
GaziESAS-Parent						
Symtoms	0.557 (<0.001)	0.410 (0.003)	0.501 (<0.001)	0.620 (<0.001)	0.729 (<0.001)	0.722 (<0.001)
Adaptive Behaviour	0.543 (<0.001)	0.173 (0.221)	0.361 (0.009)	0.465 (<0.001)	0.765(<0.001)	0.664 (<0.001)
Total	0.659 (<0.001)	0.361 (0.009)	0.526 (<0.001)	0.619 (<0.001)	0.855 (<0.001)	0.792 (<0.001)
KINDL- Patient						
Total Score	-0.410 (0.003)	-0.337 (0.015)	-0.389 (0.004)	-0.396 (0.004)	-0.316 (0.022)	-0.365 (0.008)
Physical welbeing	-0.404 (0.003)	-0.363 (0.008)	-0.403 (0.003)	-0.248 (0.076)	-0.227 (0.106)	-0.249 (0.075)
Emotional welbeing	-0.377 (0.006)	-0.270 (0.053)	-0.315 (0.023)	-0.166 (0.241)	-0.154 (0.277)	-0.166 (0.239)
Self esteem	-0.081 (0.568)	-0.098 (0.491)	-0.118 (0.406)	-0.079 (0.579)	-0.101 (0.474)	-0.101 (0.475)
Family	-0.261 (0.062)	-0.143 (0.312)	-0.220 (0.117)	-0.245 (0.081)	-0.176 (0.212)	-0.203 (0.149)
Friends	-0.241 (0.085)	0.162 (0.251)	-0.192 (0.172)	-0.107 (0.448)	-0.115 (0.419)	-0.133 (0.348)
School	-0.184 (0.190)	0.187 (0.190)	-0.200 (0.155)	-0.208 (0.139)	-0.139 (0.326)	-0.197 (0.162)
Disease	-0.111 (0.432)	-0.242 (0.084)	-0.197 (0.161)	-0.294 (0.034)	-0.123 (0.384)	-0.189 (0.180)
KINDL- Parent						
Total Score	-0.356 (0.010)	-0.167 (0.237)	-0.231 (0.099)	-0.329 (0.003)	-0.413 (<0.001)	-0.396 (<0.001)
Physical welbeing	-0.336 (0.015)	-0.272 (0.051)	-0.294 (0.035)	-0.359 (0.001)	-0.471 (<0.001)	-0.452 (<0.001)
Emotional welbeing	-0.377 (0.006)	-0.153 (0.278)	-0.239 (0.088)	-0.144 (0.209)	-0.228 (0.044)	-0.205 (0.072)
Self esteem	-0.053 (0.709)	0.157 (0.268)	0.076 (0.594)	0.040 (0.727)	0.014 (0.906)	0.028 (0.806)
Family	-0.240 (0.087)	-0.008 (0.957)	-0.107 (0.452)	-0.274 (0.015)	-0.320 (0.004)	-0.316 (0.005)
Friends	-0.077 (0.586)	0.034 (0.813)	0.002 (0.988)	-0.110 (0.340)	-0.048 (0.679)	-0.078 (0.496)
School	-0.122 (0.389)	-0.132 (0.352)	-0.131 (0.356)	-0.237 (0.036)	-0.144 (0.207)	-0.193 (0.090)
Disease	-0.092 (0.518)	-0.128 (0.365)	-0.085 (0.551)	-0.249 (0.028)	-0.273 (0.016)	-0.268 (0.017)
Duration of EoE, months	-0.297 (0.032)	-0.113 (0.423)	-0.213 (0.129)	-0.325 (0.003)	-0.273 (0.012)	-0.317 (0.003)

r: Spearman's Correlations Coefficient

Table 2. Discriminant validity testing: comparing the Tr-PEESS v2.0 and GaziESAS Domain Scores among EoE treatment groups

		Off treatment† (G1)	Under treatment (G2)	Uncomplaint with treatment (G3)	*P**
Tr-PEESS v2.0 Children and Teens Report	Frequency Domain Median (min-max)	9.1 (2.3-40.9)	11.4 (0-29.6)	13.6 (0-50)	G1 vs. G2= 0.809 G1 vs. G3=0.109 G2 vs. G3=0.285
	Severity Domain Median (min-max)	8.3 (0-38.9)	11.1 (0-55.6)	11.1 (0-61.1)	G1 vs. G2=0.701 G1 vs. G3=0.183 G2 vs. G3=0.499
	Total Median (min-max)	12.5 (1.3-37.5)	13.8 (0-40)	11.3 (0-50)	G1 vs. G2= 0.732 G1 vs. G3=0.146 G2 vs. G3=0.546
Tr-PEESS v2.0 Parent Report	Frequency Domain Median (min-max)	2.3 (0-47.7)	7.9 (0-54.5)	9.1 (0-68.2)	G1 vs. G2= 0.012 G1 vs. G3=0.010 G2 vs. G3=0.483
	Severity Domain Median (min-max)	12.5 (0-83.3)	23.6 (0-88.9)	27.8 (0-77.8)	G1 vs. G2= 0.037 G1 vs. G3=0.020 G2 vs. G3=0.374
	Total Median (min-max)	7.5 (0-63.8)	16.8 (0-70)	18.8 (0-72.5)	G1 vs. G2= 0.013 G1 vs. G3=0.007 G2 vs. G3=0.352
GaziESAS-Parent					
	Symtoms Median (min-max)	6.6 (0-65.6)	7.9 (0-65.8)	17.1 (0-57.9)	G1 vs. G2= 0.682 G1 vs. G3=0.031 G2 vs. G3=0.039
	Adaptive Behaviour Median (min-max)	10.4 (0-95.8)	20.8 (0-91.7)	27.1 (0-79.2)	G1 vs. G2= 0.111 G1 vs. G3=0.007 G2 vs. G3=0.601
	Total Median (min-max)	8.1 (0-77.4)	15.3 (0-75.8)	23.4 (1.6-53.2)	G1 vs. G2= 0.167 G1 vs. G3=0.005 G2 vs. G3=0.129

**Mann-Whitney u Test.*

† Off treatment: It refers to children who are in remission of EoE without any EoE treatment for at least a year and children who are under evaluation for remission without any EoE treatment.

Reliability of Tr-PEESS v2.0

Cronbach's α values of Tr-PEESS v2.0 ranged from 0.614 to 0. 895 (Table 3). Item-total correlation values of items 1, 2, 15, 19 (in the children and teens report), and item 1 (in the parent report) were below the desired value of 0.2. Removal of these items did not significantly increase the Cronbach's alpha values of the scales. The ICC values for Tr-PEESS v2.0 and individual domains ranged from 0.646 to 0.910, a good sign of test–retest reliability (Table 3).

Score Distribution and the Ceiling and Floor Effects of Tr-PEESS v2.0

The results of the descriptive statistics of the parent and child report scores are presented in Table 4. There were no floor or ceiling effects for total and individual domains' score of Tr-PEESS v2.0 (<20%).

DISCUSSION

This study is the first step taken to assess Tr-PEESS v2.0 as one of the PRO of EoE in Turkish children and comprises the largest sample size among similar validation studies about PEESS v2.0.[1-3] It demonstrated the reliability, convergent, and discriminant validity of Tr-PEESS v2.0, which have not studied for original PEESS v2.0, yet. Within this study, a new scale for children with EoE was developed: GaziESAS. GaziESAS is a parent-proxy scale that conforms to all scale developmental steps that assess children's disease-specific symptoms and adaptive behaviors with EoE in the last month.

Table 3. Reliability Statistics for Individual Domains Scores of Tr-PEESS v2.0

				Test–Retest Reliability	
		Cronbach's α	Item-Total Correlation Range (Min–Max)	ICC*	*P*
Children and Teens Report	Frequency domain	0.614	0.165-0.432[a]	0.811 (0.670-0.891)	**<.001**
	Severity domain	0.754	0.177-0.646[b]	0.646 (0.383-0.797)	**<.001**
	Total	0.823	0.133-0.620[c]	0.855 (0.748-0.917)	**<.001**
Parent Report	Frequency domain	0.767	0.147-0.671[d]	0.817 (0.717-0.881)	**<.001**
	Severity domain	0.853	0.374-0.701	0.910 (0.862-0.942)	**<.001**
	Total	0.895	0.264-0.727	0.876 (0.809-0.920)	**<.001**

ICC, Intraclass correlation coefficients (95% CI).
[a]Item 1 = 0.199 (Cronbach's α if item deleted = 0.607), item 15 = 0.165 (Cronbach's α if item deleted = 0.609), item 19 = 0.197 (Cronbach's α if item deleted = 0.619).
[b]Item 2 = 0.177 (Cronbach's α if item deleted = 0.765).
[c]Item 19 = 0.133 (Cronbach's α if item deleted = 0.835).
[d]Item 1 = 0.147 (Cronbach's α if item deleted = 0.771).

Table 4. Descriptive Statistics on Individual Domain Scores and Total Scores of Tr-PEESS v2.0

		Mean ± SD	Median (Min–Max)	% With Floor Effect	% With Ceiling Effect
Children and Teens Report	Frequency domain score	15.9 ± 11.8	11.4 (0-50)	3.8	0
	Severity domain score	16.9 ± 15.9	11.1 (0-61.1)	19.2	0
	Total score	16.4 ± 12.7	13.8 (0-50)	3.8	0
Parent Report	Frequency domain score	13.9 ± 16.3	6.8 (0-68.2)	17.9	0
	Severity domain score	25.5 ± 20.4	19.4 (0-88.9)	7.1	0
	Total score	18.6 ± 17.3	12.5 (0-72.5)	7.1	0

There has been only one report concerning the translation and cultural adaptation of PEESS v2.0 by using only linguistic validation in another language.[23] In that study Santos et al. pretested only 5 patients with EoE and 10 parents. The present study is the second validation study of PEESS v2.0 in another language. However, our study is not a linguistic validation study. Testing only the linguistic validity of the original scale is not an adequate approach for safe implementation in the target group of patients in that language. It is necessary to test the validity and reliability with several different methods simultaneously.[19,24,25] PEESS v2.0 was developed with 24 children and 51 parents.[3] In the second study aimed to test CsV of PEESS v2.0; they interviewed 28 patients and 46 parents.[4] Our study was conducted with 52 children (8-18 years old) and 84 parents, which makes up the largest sample size studied for this purpose.

We tested the CgV of Tr-PEESS v2.0 by 2 different scales: GaziESAS and KINDL. Although both parent and children/teens reports of Tr-PEESS v2.0 correlated well with GaziESAS, the correlation was better with Tr-PEESSv2.0 parent as expected. Both parent and children/teens report of Tr-PEESS v2.0 demonstrated moderate correlations with KINDL-patient questionnaire. Considering the disease subscale of KINDL, the correlations were present with parent forms of Tr-PEESSv2.0 and KINDL but not with Tr-PEESS v2.0 children/teens report and KINDL-patient. This finding may depreciate the CgV of children/teens report of Tr-PEESS v2.0, which covers a wide age range (8-18 years old). We think that this may be due to differences in cognitive functions and self-expressions of children from teens.

In DV analysis, Tr-PEESS v2.0 parent report was able to differentiate between children uncompliant with treatment

from children off treatment (due to remission) and under treatment, but children/teens report was not. This finding also suggests that children/teens report of PEESS v2.0 may need changes to improve.

For reliability of Tr-PEESS v2.0, the present study showed that it has satisfactory internal consistency except frequency domain of children/teens report which has a Cronbach's α value of 0.614 (acceptable borderline value). For children/teens report, the item-total correlations of items 1, 15, 19 were <0.2 for the frequency domain which is below the acceptable range.[20] The item-total correlations for the severity domain of item 2 which asks the severity of item 1 was <0.2, too. Also, in Brazil study of PEESS v2.0,[23] the authors reported that the items 15, 19, 20 for child and parent reports were not comprehensively understood and suggested changes in them. In the original developmental study of PEESS v2.0, no item analyses (item-total and item-item correlations) were presented. Therefore, it is not clear whether these items are problematic in the children/teens report of the original study or in the Turkish version. The Cronbach's α for children/teens report of Tr-PEESS v2.0 did not improve after removal of these items. Therefore we decided to keep these items to preserve the original PEESS v2.0. We believe that this issue can only be clarified with PEESSv2.0 developers in future studies.

For test–retest reliability, ICCs have not been reported yet either in the original or Brazillian versions of PEESS v2.0.[3,23] Strong test–retest reliability was identified by high ICCs>0.75 in Tr-PEESS v2.0 except for severity domain of children/teens report, which is within an acceptable range (Table 2). We do not expect high ICCs for children as parents due to differences between cognitive functions. However, this finding may be related to the time interval of retest or with a wide age range of children covered in children/teens report.

In this study, we reported for the first time floor and ceiling effect of Tr-PEESS v2.0 and GaziESAS. Lack of floor and ceiling effect confirms CnV and reliability of both scales.

Not all valid scales are reliable, and not all reliable scales are valid. Therefore, it is neccessary to test the reliability and validitiy of the scale together to develop a high quality scale. Perhaps, the greatest challenge is to come up with a high quality scale that is psychometrically sound and effective for use in research and clinical settings. In this study, to ensure that Tr-PEESS v2.0 and GaziESAS are psychometrically sound, we used a number of analytic methods to assess the validity and reliability of the scales. Content validity means "evaluation of each item to determine whether they represent the domain of interest."[24] PEESS v2.0 is a content valid scale that CsV has not been fully shown yet.[3,4] Although the developers published an article in 2015 to test CsV of PEESS v2.0, it was tested only by histological and molecular correlates of key clinical features of EoE.[4] The CsV of a scale is related to the quality of analytic methods used, such as factorial analysis, CgV, DV, item-total and item-item correlations, and ceiling and floor effects.[19,24,25]

GaziESAS showed a good performance both for CnV (CVI > 0.8) and internal consistency for reliability. It showed moderate to good correlations with the total-disease dimension of KINDL and good to excellent correlation with Tr-PEESS v2.0 for parent forms as CgV. GaziESAS has also good DV, differentiating between children uncompliant with treatment from children off treatment due to remission.

Adaptive behaviors probably arise as a compensatory mechanism to overcome dysphagia and may be used as PRO in EoE studies. In PEESS v2.0, there are only 2 items (item19,20) about adaptive behaviors within the frequency domain. However, GaziESAS has a new adaptive behavior domain that consisted of 12 items with very good analytic results to use as a PRO.

Limitations

This study has some limitations. First, the sample size of our study is enough to show reliability of the scale. However, it is not enough to test the factorial structure of the scale that is an important component of the validation. Also, there has still no study that tested the factorial structure of the original PEESS v2.0, yet. However, we tested the validity of Tr-PEESS v2.0 and GaziESAS with DVand CgV analysis that have widespread use in scale validity statistics. Our sample size is enough and suitable for these analysis. Furthermore, EoE is classified under rare diseases therefore multicenter studies are neccessary to reach larger sample sizes. Second, it is performed in Turkish children with Tr-PEESS v2.0 at a single reference center which may limit the generalizability of the results. Third, GaziESAS is a parent-proxy scale as yet waiting for development of children and adolescent reports and severity domain.

CONCLUSION

Tr-PEESS v2.0 is a valid and reliable tool to use as PRO in EoE studies. GaziESAS, which was developed in this study, is a high-quality parent-proxy scale with an additional adaptive behavior domain that passed successfully for all scale developmental stages for Turkish children with EoE. Whether GaziESAS produces the same effect in other languages is to be tested in future studies.

Ethics Committee Approval: This study was approved by Gazi University Ethics Committee (IRB No: 2019-312).

Informed Consent: A written informed consent form was obtained from all parents and children ≥12 years old.

Peer Review: Externally peer-reviewed.

Author Contributions: Concept – D.Y., H.I.E.K., A.B.; Design - D.Y., H.I.E.K., A.B.; Supervision - M.N.I., B.D., A.B.; Resource - H.I.E.K.,O.E.G, S.S., B.D., A.B.; Materials - H.I.E.K.,O.E.G, S.S., B.D., A.B.; Data Collection and/or Processing - D.Y., H.I.E.K.; Analysis and/or Interpretation - D.Y., H.I.E.K., A.B.; Literature Search - D.Y., H.I.E.K., A.B.; Writing - D.Y., H.I.E.K., A.B.; Critical Reviews - O.E.G, S.S., B.D., M.N.I., D.Y., H.I.E.K., A.B.

Acknowledgment: We thank all of our patients and their families for their support to this study. We thank Prof. Dr. Ismail Karakaya and Dr. Fuat Elkonca from the Department of Measurement and Evaluation in Education, Gazi University (Ankara, Turkey) for technical support on scale development and adaptation.

Conflicts of Interest: The authors have declared that no conflicts of interest exist.

Financial Disclosure: This work has no financial support.

REFERENCES

1. Munoz-Persy M, Lucendo AJ. Treatment of eosinophilic esophagitis in the pediatric patient: an evidence-based approach. Eur J Pediatr. 2018;177(5):649-663. [CrossRef].
2. Hirano I, Furuta GT. Approaches and challenges to management of pediatric and adult patients with eosinophilic esophagitis. Gastroenterology. 2020;158(4):840-851. [CrossRef].
3. Franciosi JP, Hommel KA, DeBrosse CW, et al. Development of a validated patient-reported symptom metric for pediatric eosinophilic esophagitis: qualitative methods. BMC Gastroenterol. 2011;11:126. [CrossRef].
4. Martin LJ, Franciosi JP, Collins MH, et al. Pediatric eosinophilic esophagitis symptom scores (PEESS v2.0) identify histologic and molecular correlates of the key clinical features of disease. J Allergy Clin Immunol. 2015;135(6):1519-28.e8. [CrossRef].
5.Lucendo AJ, Molina-Infante J, Arias Á, et al. Guidelines on eosinophilic esophagitis: evidence-based statements and recommendations for diagnosis and management in children and adults.U Eur Gastroenterol J. 2017 ;5(3):335-358. [CrossRef].
6. Mapi research trust. France: PEESS v2.0 contact information and permission to use. http://www.mapitrust.org; 2020.
7.Ravens-Sieberer U, Erhart M, Wille N, et al. Generic health-related quality-of-life assessment in children and adolescents: methodological considerations. Pharmacoeconomics. 2006;24(12):1199-1220. [CrossRef].
8. Eser E, Yüksel H, Baydur H, et al. The psychometric properties of the new Turkish generic health-related quality of life questionnaire for children (Kid-KINDL). Turk Psikiyatri Derg. 2008;19(4):409-417. PMID: 19110983
9. Ebadi A, Zarshenas L, Rakhshan M, et al. Fundamentals of Scale Validation in Health Sciences. Tehran, Iran: Jame'e Negar; 2016 (in Persian).
10. Lawshe CH. A quantitative approach to content validity. Pers Psychol. 1975;28(4):563-575. [CrossRef].
11. Polit DF, Beck CT, Owen SV. : Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health. 2007;30(4):459-467. [CrossRef].
12. Davis LL. Instrument review: getting the most from your panel of experts. Appl Nurs Res ANR. 1992;5(4):194-197. [CrossRef].
13. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231-240. [CrossRef].
14. Gregory R. J. Psychology Testing: History, Principles, and Applications. 5th ed; 2007. Boston: Ally & Bacon. (w).
15. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. Anesth Analg. 2018;126(5):1763-1768. [CrossRef].
16. Meirte J, Van Daele U, Maertens K, et al. Convergent and discriminant validity of quality of life measures used in burn populations. Burns. 2017;43(1):84-92. [CrossRef].
17. Tyack Z, Simons M, Spinks A, Wasiak J. A systematic review of the quality of burn scar rating scales for clinical and research use. Burns. 2012;38(1):6-18. [CrossRef].
18. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res. 1995;4(4):293-307. [CrossRef].
19. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42. [CrossRef].
20. Johnson R.L., Morgan G.B. Survey Scales: Investigating Scale Quality. New York, NY: The Guilford Press; 2016.
21. Carlstedt E, Lexell EM, Pessah-Rasmussen H, Iwarsson S. Psychometric properties of the Swedish version of the Generel Self-Efficacy Scale in stroke survivor. Int J Rehabil Res. 2015;38(4):333-337. [CrossRef].
22. Voigt-Radloff S, Leonhart R, Schützwohl M, et al. Dementia quality of life instrument--construct and concurrent validity in patients with mild to moderate dementia. Eur J Neurol. 2012;19(3):376-384. [CrossRef].
23. Santos MFO, Barros CP, Silva CHMD, Paro HBMDS. Translation and cultural adaptation of the Pediatric Eosinophilic esophagitis Symptom Score (PEESS v2.0). J Pediatr (Rio J). 2018;94(6):642-651. [CrossRef].
24. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: A primer. Front Public Health. 2018;6:149. [CrossRef].
25. Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Saudi J Anaesth. 2017;11(Suppl 1)(suppl 1):S80-S89. [CrossRef].

S1. Psychometric evaluation of GaziESAS

Content and Construct validity of GaziESAS

Fifteen of 52 items whose CVRcritical value was <0.75 were excluded from the scale. The CnV of the remaining 37 items in GaziESAS was evaluated by CVR (0.87-1 for each item) and CVI (0.96 for total scale). Four items with insufficient item analysis and two items with the same answer "never" by all participants were omitted leaving 31 items in the final GaziESAS (Supplement 2). There was a good correlation between two domains of GaziESAS (r = 0.590, *P*< 0.001). The total score of GaziESAS was excellently correlated with the symptoms (r = 0.868, *P*< 0.001) and adaptive behaviour (r = 0.902, *P*< 0.001) domains.

In the CgV analysis, GaziESAS and KINDL for parents were applied to parents simultaneously. KINDL total score was negatively correlated with GaziESAS's symptoms (r = -0.420, *P*< 0.001), and adaptive behaviour (r = -0.343, *P*= 0.002) domains and total score (r = -0.451, *P*<0.001). Disease dimension of KINDL was negatively correlated to GaziESAS Symptoms domain (r = -0.304, p=0.007) and to GaziESAS total score (r = -0.278, *P*= 0.014), respectively.

Another supporting evidence for the CsV of GaziESAS, is negative correlation of EoE duration with symptoms (r = -0.219, *P* = 0.045) and adaptive behaviour (r = -0.294, *P* = 0.007) domains and total score (r = -0.292, *P* = 0.007).

Discriminant validity of GaziESAS

This section was given together with DV of Tr-PEESS v2.0

Reliability of GaziESAS

GaziESAS as total and as individual domains were found to have a high internal consistency with Cronbach's α 0.895 for the total score, 0.843 for the symptoms and 0.848 for the adaptive behaviors domain, respectively. Item-total score CCs were 0.230 to 0.628. Item-symptoms domain score and item-adaptive behaviour domain score CCs were 0.282 to 0.558 and 0.301 to 0.693, respectively.

Scores and the ceiling and floor effects of GaziESAS

GaziESAS consisted of 2 domains with 31 items (19 items in symptoms domain and 12 items in adaptive behaviors domain). The range for total and domain scores of GaziESAS is 0-100, with a higher score indicating worse disease. The frequencies of minimum and maximum scores of GaziESAS were 13.1% and 0% (less than 20%), respectively according to the floor and ceiling effects analyses.

SUPPLEMENT 2. GAZIESAS

Gazi University **E**osinophilic Esophagitis and **A**daptive Behaviour **S**cale**: GaziESAS**

Lütfen aşağıda belirtilen Eozinofilik Özofajit (EÖ) ile ilgili sorunları ve davranışları dikkatlice okuyunuz. Son 1 ayda çocuğunuzun durumuna göre en uygun seçeneği işaretleyiniz.

	Hiçbir zaman	*Bazen*	*Sık sık*
A) Semptomlar			
1. Kusma	⓪	①	②
2. Karın ağrısı	⓪	①	②
3. İştahsızlık	⓪	①	②
4. Öğürme	⓪	①	②
5. Abartılı /aşırı geğirme	⓪	①	②
6. Bulantı	⓪	①	②
7. Midede ağrı	⓪	①	②
8. Göğüsde yanma-ağrı	⓪	①	②
9. Ağız kokusu	⓪	①	②
10. Ağzına acı-ekşi su gelmesi	⓪	①	②
11. Yediklerinin ağzına gelmesi	⓪	①	②
12. Sert/katı gıdaları yutamama	⓪	①	②
13. Yemeği yutarken zorlanma	⓪	①	②
14. Yemek yerken tekrar tekrar yutkunma	⓪	①	②
15. Besinin boğazına takıldığını hissetme	⓪	①	②
16. Besinin boğaza takılması ve yutamama	⓪	①	②
17. Besin takılınca kusup rahatlama	⓪	①	②
18. Kilo alamama	⓪	①	②
19. Su, süt, meyve suyu gibi sıvı besinlerle öksürük	⓪	①	②
B) Adaptif Davranışlar Çocuk ve ailesinin geliştirdiği adaptif davranışlar			
20. Yemeği yaşıtlarına göre ağır-yavaş yeme	⓪	①	②
21. Yemek yerken çok çiğneme	⓪	①	②
22. Yemek yerken çok su içme	⓪	①	②
23. Yemeği küçük lokmalara bölerek yeme	⓪	①	②
24. Yemeği ağızda bekletme	⓪	①	②
25. Çabuk/erken doyma	⓪	①	②
26. Yumuşak/blenderdan geçirilmiş besinleri tercih etme	⓪	①	②
27. Sert/katı/yumuşak olmayan besinleri yemekten kaçınma/isteksizlik	⓪	①	②
28. Besinleri yumuşatmak için karıştırarak yeme (pilavı yoğurtla karıştırmak gibi)	⓪	①	②
29. Yemek saatlerinde ağlama/mızmızlık/huysuzluk	⓪	①	②
30. Kusmasın diye daha yumuşak ve sulu yemekleri pişirmek	⓪	①	②
31. Yemek yedirmek için ısrarcı olmak/zorlamak	⓪	①	②