To cite this article: Taflan MG, Akdeniz S, Yilmaz M, Selçuk Kusderci H, Bulbul SS, Dogru S, Suren M. Validation and psychometric evaluation of the Turkish version of the Pediatric Anesthesia Emergence Delirium Scale. Turk J Clin Lab. 2026;1:1864192.

Research Article

Validation and psychometric evaluation of the Turkish version of the Pediatric Anesthesia Emergence Delirium Scale

Pediatrik Anestezi Uyanma Deliryum Ölçeği'nin Türkçe versiyonunun geçerlilik ve psikometrik değerlendirmesi

Mehmet Gokhan Taflan[1]*, Sevda Akdeniz[1], Mert Yilmaz[1], Hatice Kusderci[1], Serap Samut Bulbul[2], Serkan Dogru[3], Mustafa Suren[1]

[1]Department of Anesthesiology and Reanimation, Samsun Training and Research Hospital, Samsun University, Samsun, Türkiye

[2]Department of Pediatric Surgery, Samsun Training and Research Hospital, Samsun University, Samsun, Türkiye

[3]Department of Anesthesiology and Reanimation, Mersin City Training and Research Hospital, Mersin, Türkiye

Abstract

Aim: To evaluate the validity, reliability, construct structure, and screening concordance of the Turkish-adapted Pediatric Anesthesia Emergence Delirium scale (PAED-T).

Material and Methods: In this single-center, prospective observational study, pediatric patients aged 2–16 years (n=364) were enrolled. The PAED scale was administered at 0 and 10 minutes. Convergent and discriminant validity were examined using the FLACC scale and the Modified Aldrete Score (MAS), respectively. Internal consistency was assessed using Cronbach's alpha and, when appropriate, ordinal omega (ω). Construct validity was evaluated with the Kaiser–Meyer–Olkin (KMO) measure, Bartlett's test, and principal component analysis with varimax rotation. Screening performance (concordance) was analyzed by ROC curves (AUC), sensitivity/specificity, Youden's J, and likelihood ratios (LR+/LR–), using FLACC>3 as a behavioral reference measure rather than a delirium-specific reference standard.

Results: Mean age was 73.4±29.8 months; 97.3% were male. Internal consistency for PAED-T was α=0.728–0.741. A two-component solution outperformed a one-component model: ED-I (eye contact, purposeful actions, awareness of surroundings; loadings 0.708–0.877; ω=0.750) and ED-II (restlessness, inconsolability; loadings 0.898–0.909; ω=0.833). PAED-T showed a moderate positive correlation with FLACC (ρ=0.373) and a weak negative correlation with MAS (ρ=–0.163). ROC analysis yielded an AUC of 0.794; at a cutoff ≥10, sensitivity was 87.5%, specificity 59.4%, and LR– 0.21.

Conclusion: PAED-T is a valid and reliable tool for screening and monitoring behaviors consistent with emergence delirium in pediatric anesthesia. A cutoff ≥10 is pragmatic for screening; however, ROC findings should be interpreted as screening concordance, given that FLACC is not delirium-specific.

Keywords: delirium, general anesthesia, emergence delirium, pediatric anesthesia, validity, ROC curve

Corresponding Author*: Mehmet Gokhan Taflan, MD. Department of Anesthesiology and Reanimation, Samsun Training and Research Hospital, Samsun University, Samsun, Türkiye.
E-mail: drmehmettaflan@gmail.com
Orcid: 0000-0003-3332-790X
Doi: 10.18663/tjcl.1864192
Received: 24.01.2026 accepted: 13.03.2026 Publication date: 28.03.2026

TJCL 2026; 1: 1864192

Öz

Amaç: Bu çalışmanın amacı, Pediatrik Anestezi Uyanma Deliryum (PAED) ölçeğinin Türkçeye uyarlanmış versiyonunun (PAED-T) geçerlilik, güvenilirlik ve tarama performansını değerlendirmektir.

Gereç ve Yöntemler: Tek merkezli, prospektif, gözlemsel çalışmaya 2–16 yaş arası pediatrik hastalar (n=364) dâhil edildi. PAED ölçeği 0. ve 10. dakikalarda uygulandı. Yakınsama ve ayırt edici geçerlilik sırasıyla FLACC ve Modifiye Aldrete Skoru (MAS) ile incelendi. İç tutarlılık Cronbach alfa ve gerekli durumlarda ordinal omega (ω) ile değerlendirildi. Yapı geçerliliği için KMO, Bartlett testi ve varimaks rotasyonlu temel bileşenler analizi kullanıldı. Tarama performansı, deliryuma özgü bir altın standart olmayıp davranışsal bir referans ölçüt (ağrı/ajitasyon göstergesi) olarak kullanılan FLACC>3'e göre ROC analizi (AUC), duyarlılık/özgüllük, Youden's J ve olasılık oranları (LR+/LR–) ile değerlendirildi.

Bulgular: Ortalama yaş 73,4±29,8 ay; hastaların %97,3'ü erkekti. PAED-T için iç tutarlılık α=0,728–0,741 bulundu. İki faktörlü çözüm tek faktörlü modele kıyasla daha iyi performans gösterdi: ED-I (göz teması, amaca yönelik hareketler, çevrenin farkında olma; yükler 0,708–0,877; ω=0,750) ve ED-II (huzursuzluk, teselli edilememesi; yükler 0,898–0,909; ω=0,833). PAED-T ile FLACC arasında orta düzey pozitif korelasyon (ρ=0,373), MAS ile zayıf negatif korelasyon saptandı (ρ=–0,163). ROC analizinde AUC=0,794 idi; ≥10 eşik değerinde duyarlılık %87,5, özgüllük %59,4 ve LR–=0,21 bulundu.

Sonuç: PAED-T, pediatrik anestezide uyanma deliryumunun taranması ve izlenmesinde geçerli ve güvenilir bir araçtır. Klinik tarama için ≥10 eşik değeri pratik görünmekle birlikte, FLACC'nin deliryuma özgü olmaması nedeniyle ROC bulguları tarama bağlamında yorumlanmalıdır.

Anahtar Kelimeler: deliryum, genel anestezi, uyanma deliryumu, pediatrik anestezi, geçerlilik, ROC eğrisi

Introduction

Emergence delirium (ED) is a transient yet clinically consequential condition in children during emergence from general anesthesia, characterized by disorientation, restlessness, purposeless motor activity, agitation, and an inability to interact with the environment [1]. Reported incidence varies widely, ranging from 2% to 80%; this variability is largely attributable to heterogeneity in definitions and assessment tools used across studies [2]. Because of its adverse impact on patient safety, parental satisfaction, and the burden on health-care systems, ED is considered a significant concern in pediatric anesthesia [3].

Numerous risk factors have been identified for the development of ED. Younger age (<6 years), male sex, high preoperative anxiety, the use of short-acting volatile agents such as sevoflurane, and otolaryngologic and ophthalmologic surgeries are among the most frequently reported factors [4]. In addition, comorbid neuropsychiatric conditions such as attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) increase the risk of ED [5]. In the short term, ED can lead to complications such as removal of intravenous lines or catheters, disruption of surgical-site integrity, and the need for additional analgesics or sedatives; in the long term, it has been associated with sequelae including sleep disturbances, separation anxiety, and behavioral problems [6].

Preventive strategies for ED include pharmacologic approaches such as propofol, dexmedetomidine, ketamine, and opioids and nonpharmacologic strategies, including parental presence during induction, video-based interventions, virtual reality, and distraction games [7]. Nonetheless, accurate diagnosis of ED requires the use of reliable and standardized measurement instruments.

Developed for this purpose, the Pediatric Anesthesia Emergence Delirium (PAED) scale is the most commonly used instrument for assessing ED in contemporary clinical practice [8]. The five-item scale is scored from 0-20, and a cutoff of ≥10 is commonly used as a pragmatic screening threshold in clinical practice. However, some studies have suggested that higher thresholds (e.g., ≥12) may reduce false positives, highlighting that the preferred cut-off may depend on the clinical context and the intended balance between sensitivity and specificity [9].

The PAED scale has been adapted into various languages across different countries, and its psychometric properties have been examined. Studies from Scandinavian countries have reported high internal consistency and substantial interrater

agreement [10]. The Spanish version has demonstrated validity and reliability [11]. In a Swedish psychometric evaluation, a one-factor structure was supported, with excellent internal consistency (Cronbach's α≈0.96) and near-perfect interrater reliability [12]. Studies conducted in Asian countries have also supported the scale's cultural adaptability [9]. These findings support the PAED as a valid and reliable measurement tool across diverse cultural and linguistic contexts.

Validation and reliability studies of the PAED scale have not yet been reported in the Turkish-speaking population. To enable standardized diagnosis of ED in clinical practice and scientific research, this gap needs to be addressed. The aim of this study is to translate the PAED scale into Turkish and conduct its cultural adaptation; to evaluate its psychometric validity and reliability; and to determine its screening performance using receiver operating characteristic (ROC) analysis.

Materials and Methods

This prospective, observational, single-center validity and reliability study was conducted at the Department of Anesthesiology and Reanimation, Samsun University Samsun Training and Research Hospital, between March 1, 2023, and September 30, 2023. Ethical approval was obtained from an institutional clinical research ethics committee. All procedures were carried out in accordance with the 2013 revision of the Declaration of Helsinki [13]. Written informed consent was obtained from the parents or legal guardians of all children included in the study.

Study population and sample size justification

Only pediatric patients aged 2–16 years were included in the study; cases outside this range were excluded. Children younger than 2 years were excluded because emergence behaviors may differ developmentally in infants and toddlers, and this age group was outside the predefined validation target population. Eligible patients consisted of children who underwent general anesthesia for elective surgery. Exclusion criteria were emergency surgeries; a history of metabolic, endocrine, anatomic, or neurologic disease; active psychiatric treatment; use of preoperative sedative medication; surgeries performed under local anesthesia only; and age <2 years. A total of 413 cases were screened; 27 were excluded because parental consent could not be obtained, and 22 due to missing data. Consequently, 364 pediatric patients were included in the analysis. Sample size was determined based on the rule of at least 5-10 participants per item recommended for validity and reliability studies, as well as the psychometric parameters reported in the relevant Spanish adaptation study [11]. These approaches support that at least 150-200 participants are sufficient for factor analysis and internal consistency assessments; in the present study, with n=364, this threshold was substantially exceeded.

Translation and Cultural Adaptation

The Turkish adaptation of the PAED scale was conducted in accordance with Brislin's five-stage translation model [14]. In the first stage, the scale was translated from English into Turkish by three independent translators. The resulting translations were evaluated by an anesthesiology specialist for conceptual coherence and clinical appropriateness. Subsequently, back-translation was performed by three different translators who were unfamiliar with the original scale. The back-translations were compared with the original form, and equivalence of meaning and content was evaluated by an expert committee composed of anesthesiologists and linguists, after which necessary revisions were made. In the final stage, the draft form was piloted in 30 pediatric patients; based on feedback regarding comprehensibility and cultural fit, revisions were made, and the final Turkish version (PAED-T) was produced.

Data Collection Procedures

Anesthesia induction and maintenance were performed in accordance with institutional standard protocols. For induction, intravenous propofol 2–5 mg/kg was administered; rocuronium 0.6-1 mg/kg was given for neuromuscular blockade. Maintenance was provided with sevoflurane targeting an age-adjusted 1.0 MAC; per institutional protocol, a 50% N_2O/O_2 mixture was used as needed, and the total alveolar anesthetic concentration of sevoflurane plus N_2O was titrated to 1.0 MAC. For intraoperative analgesia, fentanyl was titrated within the pediatric dosing range with a 1-2 µg/kg bolus at induction and additional 0.5-1 µg/kg boluses as required. All patients underwent anesthesia monitoring in accordance with ASA standards (electrocardiography, noninvasive blood pressure measurement, pulse oximetry, and capnography), and hemodynamic parameters were recorded regularly throughout the operation.

Emergence delirium was assessed using the Pediatric

Anesthesia Emergence Delirium (PAED) scale. The scale was administered at two time points: (i) the moment when the child opened the eyes, resumed spontaneous breathing, and first engaged with the environment (0 minutes), and (ii) 10 minutes after emergence. Each assessment was performed by an experienced anesthesiologist (S.A.) who was blinded to the study hypotheses. All PAED-T, FLACC, and MAS ratings at both time points were performed by the same assessor to ensure consistency across assessments. To test convergent validity, two additional scales were administered at the same time points: Face, Legs, Activity, Cry, and Consolability (FLACC) and the Modified Aldrete Score (MAS).

Statistical Analysis

All analyses were performed using IBM SPSS Statistics, version 20.0 (IBM Corp., Armonk, NY). Continuous variables were presented as mean ± standard deviation (SD) or median [IQR], and categorical variables as n (%). Internal consistency was evaluated for total and subscales using Cronbach's alpha (α) and, where appropriate, ordinal omega (ω); item–total and inter-item relationships were reported with Spearman's rho (ρ). For construct validity, sample adequacy was confirmed with KMO = 0.608 and Bartlett's test of sphericity ($p<0.001$); the component structure was examined using principal component analysis (PCA) extracting components with eigenvalues >1 and varimax rotation (loading threshold ≥0.50). Because PAED items are ordinal, PCA was used as an exploratory data-reduction approach; nevertheless, factor-analytic methods based on polychoric correlations may better represent latent constructs in ordinal scales and should be prioritized in future confirmatory studies. Convergent and discriminant validity were tested by correlations among PAED-T, FLACC, and MAS. Screening performance was evaluated with ROC analysis; AUC, sensitivity, and specificity, together with Youden's J, balanced accuracy, and likelihood ratios (LR+, LR–), were reported. Statistical significance was set at two-tailed $p<0.05$.

Results

Initially, 413 pediatric patients were assessed for eligibility. Twenty-seven families declined participation, and 386 patients were enrolled. During data evaluation, 22 patients were excluded due to missing data. Consequently, 364 patients were included in the statistical analyses (Figure 1).

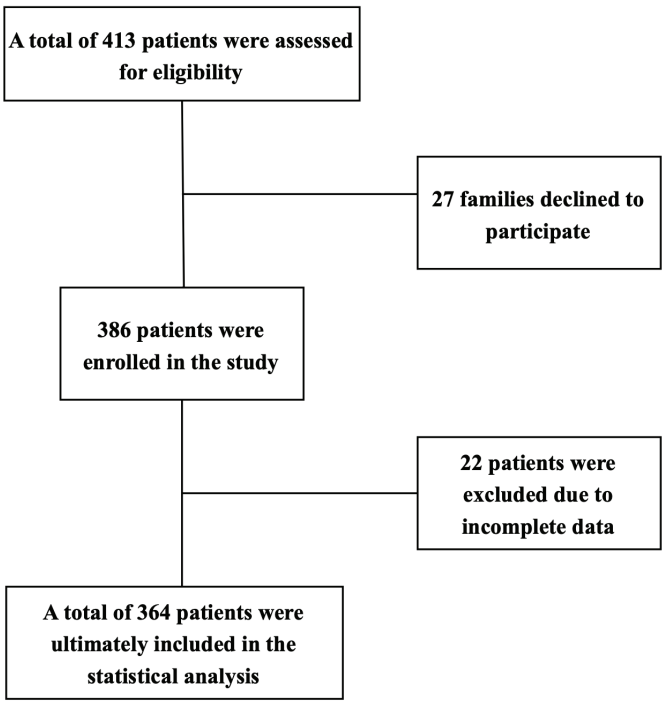

Figure 1. Flow diagram of patient recruitment and inclusion process.

Demographic and Clinical Characteristics

The mean age of participants was 73.4 ± 29.8 months (≈ 6.1 ± 2.5 years; range, 24–192 months); 97.3% were boys (n = 354) and 2.7% were girls (n = 10). The mean body mass index was 17.4 ± 3.2 kg/m^2, the mean age of the guardian was 35.3 ± 6.3 years, and the mean number of siblings was 2.5 ± 1.8. According to the ASA classification, 87.9% (n = 320) were ASA I and 12.1% (n = 44) were ASA II. A history of allergy was present in 43 patients (11.8%), and a history of previous surgery in 28 (7.7%). The mean operation time was 16.9 ± 8.8 minutes (4–89). Anesthesia combinations were as follows: propofol + sevoflurane (n = 307; 84.3%), propofol + sevoflurane + N_2O (n = 47; 12.9%), and propofol + sevoflurane + rocuronium + fentanyl (n = 10; 2.7%) (Table 1). The cohort predominantly comprised short-duration minor elective procedures, including male-predominant operations (e.g., circumcision-related urologic surgery).

Internal Consistency Reliability

The internal consistency of the PAED-T was evaluated using Cronbach's alpha (α). Overall alpha coefficients at the two time points ranged from 0.728 to 0.741; removing any single item did not meaningfully increase this value. Item–total correlations ranged from 0.631 to 0.691. When each item was deleted, the resulting alpha values remained between 0.721 and 0.741. These findings support the reliability of the Turkish version of the scale (Table 2).

Inter-item Correlations

Spearman correlation coefficients among PAED-T items ranged from ρ = 0.101 to 0.709. The strongest correlation was observed between "restlessness" and "inconsolability" (ρ = 0.709; p<0.05). Significant associations were also noted among items in the cognitive/interaction domain: "eye contact"– "purposeful actions" (ρ = 0.630; p<0.05) and "purposeful actions"– "awareness of surroundings" (ρ = 0.485; p<0.05). Correlations between "awareness of surroundings" and "restlessness" (ρ = 0.171; p = 0.001) and between "awareness of surroundings" and "inconsolability" (ρ = 0.149; p = 0.004) were weaker but remained statistically significant (Table 3).

Convergent and Discriminant Validity

A moderate positive correlation was observed between PAED-T and FLACC (ρ = 0.373; p<0.05). At the subscale level, ED-I correlated with FLACC at ρ = 0.357 (p<0.05), and ED-II at ρ = 0.219 (p<0.001). In the discriminant validity analysis, a weak but significant negative correlation was found between PAED-T and MAS (ρ = –0.163; p = 0.002). These findings support the convergent and discriminant validity of the PAED-T (Table 4).

Construct Validity and Component Structure (Principal Component Analysis, PCA)

Construct validity of the PAED-T was explored by examining its component structure using principal component analysis (PCA). Sampling adequacy was confirmed with KMO = 0.608 and Bartlett's test of sphericity (χ^2 = 515.208; p<0.001). The KMO value was at the lower acceptable bound. In the one-component solution, loadings ranged from 0.565 to 0.785, with 44.7% of the variance explained (Cronbach's α = 0.680). The two-component solution Component 1 (ED-I: eye contact, purposeful actions, awareness of surroundings) and Component 2 (ED-II: restlessness, inconsolability) showed better psychometric performance (ED-I loadings 0.708–0.877; ED-II loadings 0.898–0.909; α = 0.735 and 0.792; ordinal ω = 0.750 and 0.833) (Table 5).

Concordance with the Behavioral Reference Measure and Screening Performance (ROC)

Using FLACC>3 as a behavioral reference measure (pain/agitation proxy), a cutoff of ≥10 was selected as a pragmatic screening threshold. At this level, sensitivity was 87.5% and specificity was 59.4%. ROC analysis yielded an AUC of 0.794 (p<0.05). Youden's index was J = 0.469, balanced accuracy was 0.735, and likelihood ratios were LR+ = 2.16 and LR– = 0.21. The ROC curve is presented in Figure 2. Because FLACC is not delirium-specific, these ROC metrics should be interpreted as screening concordance with pain/agitation-related behaviors rather than delirium-specific diagnostic accuracy.

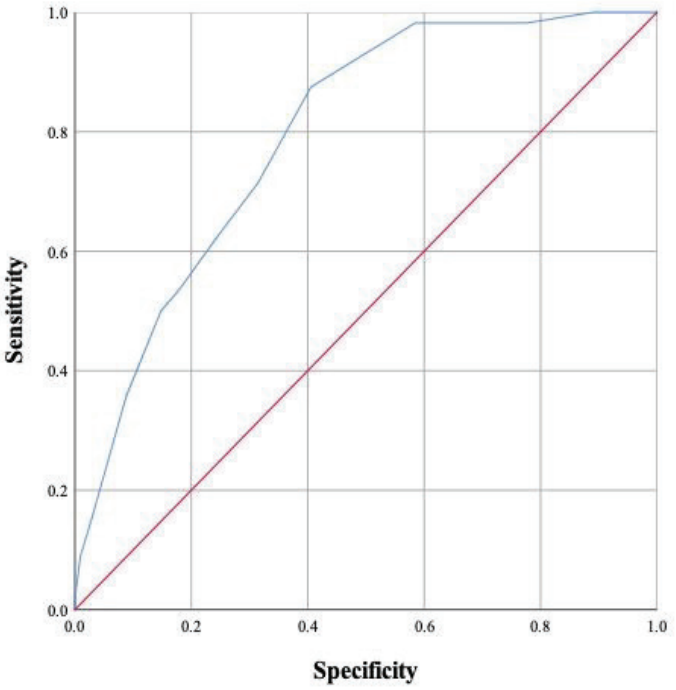

Figure 2. ROC curve for PAED-T against FLACC > 3 (reference standard). AUC = 0.794 (p<0.050). At PAED-T ≥ 10: sensitivity 87.5%, specificity 59.4% (n = 364). Abbreviations: PAED-T, Turkish version of the Pediatric Anesthesia Emergence Delirium scale; FLACC, Face, Legs, Activity, Cry, and Consolability.

Table 1. Demographic and clinical characteristics of the study population (n = 364).

	N (%)	Mean ± SD	Range (Min–Max)
Age (months)		73.4 ± 29.8	24 – 192
Gender			
Boys	354 (97.3)		
Girls	10 (2.7)		
BMI (kg/m2)		17.4 ± 3.2	11.5 – 31.4
Guardian age (years)		35.3 ± 6.3	19 – 52
Number of siblings			
1	54 (14.4)		
2	167 (44.7)		
3	114 (30.5)		
4	30 (8)		
5	6 (1.6)		
8	3 (0.8)		
ASA			
I	320 (87.9)		
II	44 (12.1)		
Presence of allergy	43 (11.8)		
History of any operation	28 (7.7)		
Operation time (min)		16.9 ± 8.8	4 – 89
Anesthesia			
Propofol + Sevoflurane	307 (84.3)		
Propofol + Sevoflurane + Nitrous oxide	47 (12.9)		
Propofol + Sevoflurane + Rocuronium + Fentanyl	10 (2.7)		

Table 2. Item-level properties of the PAED-T scale.

	Median (Q1–Q3)	Item score distribution					ITC[b]	Alpha[c]
		0	1	2	3	4		
Eye contact[a]	3 (2, 3)	16.3	40.4	36.4	7	—	0.631	0.741
Purposeful actions[a]	3 (2, 3)	16	36.6	40.9	6.1	0.3	0.691	0.728
Aware of surroundings[a]	3 (2, 3)	19.3	35.6	33.2	11.8	0.3	0.647	0.734
Restless	1 (0, 2)	27.5	39.8	21.7	7.8	3.2	0.632	0.727
Inconsolable	1 (0, 2)	42.2	30.2	17.9	7.2	2.4	0.639	0.721

0=Not at all, 1=Just a little, 2=Quite a bit, 3=Very much, 4=Extremely.
[a] Reversed scored items.
[b] Item-total correlations adjusted for overlaps.
[c] Cronbach's alpha value if item deleted.

Table 3. Inter-item correlation matrix of the PAED T.

	Eye con-tact	Purposeful actions	Aware of surroundings	Restless	Inconsolable
Eye contact					
Purposeful actions	ρ = 0.630, p < 0.05				
Aware of surroundings	ρ = 0.352, p < 0.05	ρ = 0.485, p < 0.05			
Restless	ρ = 0.101, p = 0.050	ρ = 0.142, p = 0.006	ρ = 0.171, p = 0.001		
Inconsolable	ρ = 0.125, p = 0.015	ρ = 0.159, p = 0.002	ρ = 0.149, p = 0.004	ρ = 0.709, p < 0.05	

Cells report Spearman's rho (ρ); only the lower triangle is displayed (diagonal shown as "—"). Two-tailed tests were used, with statistical significance defined as p<0.050. p-values are reported to three decimals; entries with p≥0.050 were considered non-significant. Abbreviation: PAED-T, Turkish version of the Pediatric Anesthesia Emergence Delirium scale.

Table 4. Convergent and discriminant validity: Correlation matrix of PAED T, FLACC, and MAS.

	PAED-T (ρ, p)	FLACC (ρ, p)	MAS (ρ, p)
PAED-T	—	—	—
FLACC	ρ = 0.373, p < 0.05	—	—
MAS	ρ = –0.163, p = 0.002	ρ = –0.262, p < 0.05	—

Cells report Spearman's rho (ρ); only the lower triangle is displayed (diagonal shown as "—"). Two-tailed tests were used; statistical significance was defined as p<0.050. Abbreviations: PAED-T, Turkish version of the Pediatric Anesthesia Emergence Delirium scale; FLACC, Face, Legs, Activity, Cry, and Consolability scale; MAS, Modified Aldrete Score.

Table 5. Construct validity: Factor analysis of the PAED T.

	One-factor model	Two-factor model	
	Factor loadings	Component 1 Factor loadings (ED I)	Component 2 Factor loadings (ED II)
Eye contact	0.697	0.832	
Purposeful actions	0.785	0.877	
Aware of surroundings	0.676	0.708	
Restless	0.565		0.909
Inconsolable	0.597		0.898
Reliability analysis			
Cronbach's α	0.680	0.735	0.792
Ordinal Ω	0.627	0.750	0.833

Factor extraction followed the Kaiser criterion (eigenvalues > 1); varimax (orthogonal) rotation was applied; factor loadings ≥ 0.50 were deemed salient. Sampling adequacy was acceptable (KMO = 0.608), and Bartlett's test of sphericity indicated sufficient inter-item correlations (χ^2 = 515.208, p < 0.001). ED-I = emergence delirium–specific behaviors (eye contact, purposeful actions, awareness of surroundings); ED-II = emergence delirium–nonspecific behaviors (restlessness, inconsolability).

Discussion

This study demonstrates that the Turkish adaptation of the Pediatric Anesthesia Emergence Delirium (PAED) scale (PAED-T) is valid and reliable for clinical use. The similarity of internal consistency across two time points (α = 0.728–0.741), together with item–total correlations (Spearman ρ = 0.631–0.691) and inter-item correlations (ρ = 0.101–0.709) within expected ranges, supports the structural integrity of the scale. Similar measurement properties have been consistently reported in adaptations to other languages [10–12], and the psychometric suitability documented for the Spanish version parallels our findings with respect to cross-cultural applicability [11].

The two-factor solution being superior to the one-factor model suggests that the perceptual–cognitive (ED-I) and behavioral agitation (ED-II) dimensions of delirium are partially separable in clinical practice [15]. This pattern is consistent with prior psychometric evaluations of the dimensional structure of the PAED [10–12]. In our study, factor loadings for ED-I were 0.708–0.877 (α = 0.735; ω = 0.750), and for ED-II were 0.898–0.909 (α = 0.792; ω = 0.833); the strong correlation between "restlessness" and "inconsolability" (Spearman ρ≈0.71) supports the clustering of behavioral items.

Findings for convergent and discriminant validity align with clinical expectations: the positive correlation between PAED-T and FLACC (ρ≈0.37) suggests that pain- and agitation-related behaviors may partially overlap with delirium signs during early emergence. The weak negative correlation with MAS (ρ≈–0.16) indicates that as delirium severity increases, recovery performance declines. Therefore, particularly in the early emergence phase, the concurrent use of the PAED-T with pain measures in clinical assessment may reduce the risk of misclassification.

ROC findings indicate that a total-score threshold of ≥10 provides a practical balance between sensitivity (87.5%) and specificity (59.4%) for screening applications. A negative likelihood ratio (LR–) of 0.21 supports the utility of this threshold for screening purposes, particularly for reducing the likelihood of ED-related behaviors in this setting. However, the moderate specificity suggests that in contexts where specificity is prioritized such as when the cost of false-positive results is high alternative thresholds (e.g., ≥12) may be considered, taking the sensitivity–specificity trade-off into account. In this context, the ≥10 threshold should be interpreted as a pragmatic screening cut-off rather than a

diagnostic boundary. When the clinical priority is to minimize unnecessary interventions triggered by false positives, a higher threshold such as ≥12 may be preferable, albeit at the expense of reduced sensitivity. Therefore, threshold selection should be guided by the intended use (screening vs. confirmation), local workflow, and the relative consequences of false-positive versus false-negative classifications.

The age distribution, surgical spectrum, and use of volatile anesthetics in our sample are consistent with contemporary risk patterns described for ED [3]. Together with current nonpharmacologic and pharmacologic approaches, standardized measurement of the PAED-T enables the development of comparable quality indicators within clinical care pathways [7,16,17].

In this study, FLACC > 3 was used as a behavioral reference measure; because FLACC is not delirium-specific, ROC metrics reflect screening concordance with pain/agitation-related behaviors rather than delirium-specific diagnostic accuracy. Construct structure was explored using PCA with Pearson correlations; given the ordinal nature of the items, factor-analytic approaches based on polychoric correlations and ordinal-appropriate confirmatory modeling should be prioritized in future studies. In addition, the KMO value (0.608) was at the lower acceptable bound; therefore, the component solution should be interpreted cautiously. Interrater and test–retest reliability were not assessed because all ratings were performed by a single rater; thus, reproducibility across observers and temporal stability could not be evaluated. Additionally, the extreme sex imbalance (97.3% male) substantially limits external validity for female children, particularly given reports of sex-related differences in ED. This imbalance is consistent with the predominantly short-duration minor elective case mix in the present cohort, including male-predominant procedures (e.g., circumcision-related urologic surgery). Relatedly, the short mean operative time indicates limited representation of major surgery or prolonged anesthesia; therefore, the performance of PAED-T in longer or more complex procedures warrants further evaluation. The study population was limited to children aged 2–16 years; therefore, the performance of PAED-T cannot be inferred for infants and toddlers younger than 2 years. Validation studies specifically designed for the <2-year age group are needed to determine the applicability of PAED-T in the youngest pediatric patients. Finally, the single-center design may limit generalizability across institutions with different perioperative pathways, PACU practices, and surgical case mixes; multicenter validation across broader procedural spectra is warranted.

In conclusion, the PAED-T is a valid and reliable instrument for screening and monitoring emergence delirium in pediatric anesthesia. A two-factor structure ED-I (perceptual–cognitive) and ED-II (behavioral agitation) was supported. Findings for convergent and discriminant validity were consistent with clinical expectations ($\rho \approx 0.37$ with FLACC; $\rho \approx -0.16$ with MAS). ROC analysis yielded AUC = 0.794; a threshold of ≥10 was found to be practical for screening, with sensitivity 87.5% and specificity 59.4%. An LR− of 0.21 supports its utility for screening purposes. Completion of multicenter validation, interrater and test–retest reliability, and studies of responsiveness and the minimal clinically important difference (MCID) will strengthen the integration of the PAED-T into routine care pathways.

Declaration of conflicting interests

The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The authors received no financial support for the research and/or authorship of this article.

Ethics approval

Samsun University Clinical Research Ethics Committee (Approval No.: SÜKAEK-2023/3-1).

Authors' contribution

MGT: Conceptualization; Data curation; Writing, original draft; Writing, review and editing. SA: Conceptualization; Methodology; Investigation (patient recruitment); Visualization; Writing, review and editing. MY: Investigation; Formal analysis; Visualization; Methodology; Writing, review and editing. HK: Methodology; Investigation; Writing, review and editing. SSB: Investigation; Writing, review and editing. SD: Conceptualization; Methodology; Formal analysis; Writing, review and editing. MS: Conceptualization; Methodology; Supervision; Writing, review and editing. All authors read and approved the final manuscript.

References

1. Zhang W, Cheng Y, Zhang L, Wei Y, Xie H, Huang J. Association between emergence delirium and brain status parameters in children undergoing general anesthesia: A prospective observational study. Paediatr Anaesth 2024; 34: 130-37.
2. Liu K, Liu C, Ulualp SO. Prevalence of emergence delirium in children undergoing tonsillectomy and adenoidectomy. Anesthesiol Res Pract 2022; 2022: 1465999.
3. Chen YC, Foster J, Wang ML, Rohmah I, Tseng YH et al. Global prevalence and risk factors of emergence delirium in pediatric patients undergoing general anesthesia: A systemic review and meta-analysis. J Pediatr Nurs 2024; 77: 74-80.
4. Mason KP. Paediatric emergence delirium: a comprehensive review and interpretation of the literature. Br J Anaesth 2017; 118: 335-43.
5. Sobol M, Sobol MK, Kowal M. Adult behavior toward the child before surgery and pediatric emergence delirium. Paediatr Anaesth 2022; 32: 43-48.
6. Dahmani S, Delivet H, Hilly J. Emergence delirium in children: an update. Curr Opin Anaesthesiol 2014; 27: 309-15.
7. Shi X, Su L, Sun Y, Ma C, Wang Z. Non-pharmacological interventions for preventing emergence delirium in children under general anesthesia: A scoping review. J Pediatr Nurs 2024; 77: e38-53.
8. Sikich N, Lerman J. Development and psychometric evaluation of the pediatric anesthesia emergence delirium scale. Anesthesiology 2004; 100: 1138-45.
9. Bong CL, Ng ASB. Evaluation of emergence delirium in Asian children using the Pediatric Anesthesia Emergence Delirium Scale. Paediatr Anaesth 2009; 19: 593-600.
10. Simonsen BY, Skovby P, Lisby M. An evaluation of the Danish version of the Pediatric Anesthesia Emergence Delirium scale. Acta Anaesthesiol Scand 2020; 64: 613-19.
11. Epulef V, Muñoz S, Alarcón AM, Vial M. Spanish version of the Pediatric Anesthesia Emergence Delirium scale: translation and cross-cultural adaptation. BMC Anesthesiol 2022; 22: 349.
12. Ringblom J, Wåhlin I, Proczkowska M. A psychometric evaluation of the Pediatric Anesthesia Emergence Delirium scale. Paediatr Anaesth 2018; 28: 332-37.
13. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013; 310: 2191-94.
14. Jones PS, Lee JW, Phillips LR, Zhang XE, Jaceldo KB. An adaptation of Brislin's translation model for cross-cultural research. Nurs Res 2001; 50: 300-4.
15. Ringblom J, Wåhlin I, Proczkowska M, Korhonen L, Årestedt K. Measurement properties of the Pediatric Anesthesia Emergence Delirium scale: A confirmatory factor analysis-based study. Paediatr Anaesth 2025; 35: 155-62.
16. Zhao D, Tian T, Jin S. Preoperative virtual reality for pediatric patients undergoing general anesthesia: A meta-analysis of randomized controlled trial. Paediatr Anaesth 2025. doi:10.1111/pan.70016.
17. Li Y, Jiang Y, Zhang L. The effect of different dosage of intranasal dexmedetomidine on preventing emergence delirium or agitation in children: A network meta-analysis of randomized controlled trials. PLoS One 2024; 19: e0304796.