

DOI: <https://doi.org/10.5554/22562087.e1166>

Colombian translation and cultural adaptation of the Pediatric Anesthesia Emergence Delirium scale

Traducción y adaptación cultural colombiana de la escala Pediatric Anesthesia Emergence Delirium

Juan Nicolás Díaz-Riaño^{a,c} ; Germán Andrés Franco-Gruntorad^a ; Juan Camilo Jerez-López^c ; María Paula Gómez^b ; Laura Cabrera-Cumplido^a ; Paula Andrea Cuartas-González^a ; Andrea Barragán-Méndez^{a,c} ; Susana Beltrán-Villegas^{a,c} ; Juan José López-Reyes^{a,c} .

^aAnesthesiology Department, Fundación Cardioinfantil, Instituto de Cardiología. Bogotá, Colombia.

^bHospital Universitario San Ignacio, Pontificia Universidad Javeriana. Bogotá, Colombia.

^cAnesthesiology and Resuscitation, School of Medicine and Health Sciences, Universidad del Rosario. Bogotá, Colombia.

Author for correspondence: Departamento de Anestesiología, Fundación Cardioinfantil - Instituto de Cardiología, Cl. 163a # 13B-60. Bogotá, Colombia.

E-mail: gfrancog@lacardio.org

What is known about this subject?

The PAED scale is used in the world within the 30 minutes following emergence from general anesthesia in order to detect emergence delirium, a multifactorial complication for which there is no established treatment. Therefore, preventing its occurrence is of the utmost importance. This condition manifests as altered emotional states, sleep-wake cycles and psychomotor activity.

This phenomenon may last up to two days, giving rise to significant morbidity, transient neurologic deficit, prolonged hospital stay and regression in neurodevelopmental milestones, if not identified within a timely period of time.

How to cite this article:

Development of the final Spanish version for the Colombian context of the PAED scale, a consistent, sufficient, clear and relevant tool for the assessment of emergence delirium in the pediatric population, adapted and validated based on the judgement of twelve experts. This initial undertaking sets the basis for future research aimed at looking into the replicability of this adaptation in pediatric post-anesthesia care in Colombia.

How to cite this article?

Díaz-Riaño N, Franco-Gruntorad GA, Jerez-López JC, Gómez MP, Cabrera-Cumplido L, Cuartas-González PA, et al. Colombian translation and cultural adaptation of the Pediatric Anesthesia Emergence Delirium scale. Colombian Journal of Anesthesiology. 2026;54:e1166.

Abstract

Introduction

Emergence delirium is a state of mental dissociation which occurs in children as they emerge from anesthesia. Its incidence ranges between 10% and 80% and is associated with neurotoxicity, prolonged hospital stay, self-harm and aggression toward others. To date, the Pediatric Anesthesia Emergence Delirium (PAED) scale is the best studied tool for diagnosing this condition.

Objective

Given the lack of a similar tool for the Colombian context, the aim of this study was to translate and culturally adapt the PAED scale in Spanish.

Methods

Prospective study with a mixed (quantitative and qualitative) methodology approach consisting of three phases. Phase 1 comprised the translation into Spanish of the PAED tool and its cultural adaptation to the Colombian context. Phase 2 consisted of expert validations. Phase 3 consisted of a pilot test using the final version of the scale, during which two trained observers used the tool during the consecutive assessment of 49 pediatric patients in a hospital post-anesthetic care unit. Parametric statistics were used to analyze the information.

Results

With moderate to near-perfect agreement, the translation and cultural adaptation process was validated by expert judgement in order to obtain the final Spanish version of the PAED scale. In the pilot test, inter-observer agreement ranged from acceptable to near-perfect (Cohen's k 0.29-0.95; p : 0.001) for the tool's assessment categories. Absolute agreement obtained for the diagnosis of emergence delirium based on the adapted PAED scale was excellent, with an ICCA (absolute agreement) of 0.82 (0.69–0.90) to 0.92 (0.846–0.96).

Conclusions

A Spanish version of the PAED scale culturally adapted to the Colombian context was validated by expert judgement and shown to be useful in the clinical setting.

Keywords

PAED; Emergence delirium; General anesthesia; Anesthesia recovery; Pediatric anesthesia; Anesthesiology.

Lea la versión en español de este artículo en www.revcolanest.com.co

Copyright © 2025 Sociedad Colombiana de Anestesiología y Reanimación (S.C.A.R.E.).

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Resumen

Introducción: El delirio de emergencia es un estado de disociación mental en niños durante el despertar anestésico. Su incidencia oscila entre 10 % y 80 %, y se asocia a neurotoxicidad, aumento de los tiempos de hospitalización, autoagresión y heteroagresión. La escala Pediatric Anesthesia Emergence Delirium (PAED) es el instrumento más estudiado en la actualidad para su diagnóstico.

Objetivo: Ante la falta de un instrumento similar para el contexto colombiano, este estudio tuvo como objetivo la traducción al español y adaptación cultural de la escala PAED.

Métodos: Estudio prospectivo con enfoque de metodología mixta (cuantitativa y cualitativa), desarrollado en tres fases. La fase 1 consistió en la traducción al español de la escala PAED y su adaptación cultural al contexto colombiano. La fase 2 correspondió a la validación por juicio de expertos. En la fase 3 se llevó a cabo una prueba piloto utilizando la versión final de la escala, en la que dos observadores capacitados aplicaron la herramienta durante la evaluación consecutiva de 49 pacientes pediátricos en la unidad de recuperación posanestésica de un centro hospitalario. La información fue analizada con estadística paramétrica.

Resultados: Con un grado de concordancia entre moderado y casi perfecto, el juicio de expertos validó el proceso de traducción al español y adaptación cultural de la escala para obtener la versión final en español de la escala PAED. En la prueba piloto se obtuvo una concordancia entre observadores aceptable hasta casi perfecta (κ de Cohen = 0,29–0,95; $p = 0,001$) para las categorías de evaluación de la escala. El acuerdo absoluto obtenido para el diagnóstico de delirio de emergencia según la escala PAED adaptada fue excelente con un CCI (acuerdo absoluto) = 0,82 (0,69–0,90) a 0,92 (0,846–0,96).

Conclusiones: Se obtuvo una versión en español adaptada culturalmente al contexto colombiano de la escala PAED, validada por juicio de expertos y con utilidad demostrada en el ámbito clínico.

Palabras clave: PAED; Delirium de emergencia; Anestesia general; Recuperación anestésica; Anestesia pediátrica; Anestesiología.

INTRODUCTION

Emergence delirium (ED) is a state of mental dissociation which manifests 10 to 30 minutes after waking from general anesthesia. It increases hospital length of stay and medical care costs, and has been associated with various risk factors (1,2). Incidence ranges between 10% and 80% in patients receiving sevoflurane anesthesia (3), which results in a 1.43-fold increase in the risk of developing behavioral changes, self-harm and heteroaggressiveness, among other complications (4,5).

Although no specific treatment has been found to date, timely prevention and diagnosis are crucial, especially in children 3 to 6 years of age (6). To this end, the PAED scale, developed in 2024, has been validated and used in countless studies as an objective and reliable tool for diagnosing this condition (7-9). However, in the Colombian clinical context there is no evidence of the use of any such specific diagnostic scale. At present, detection is based on clinical assessment — subjective

and heterogeneous — which explains the variation in prevalence found in the literature and suggests the need for an objective and reliable tool. The aim of this study was to undertake the Spanish translation and cultural adaptation (TCA) of the PAED tool to facilitate its use in the clinical context of Colombia and contribute to timely and reliable diagnosis of emergence delirium in the pediatric population.

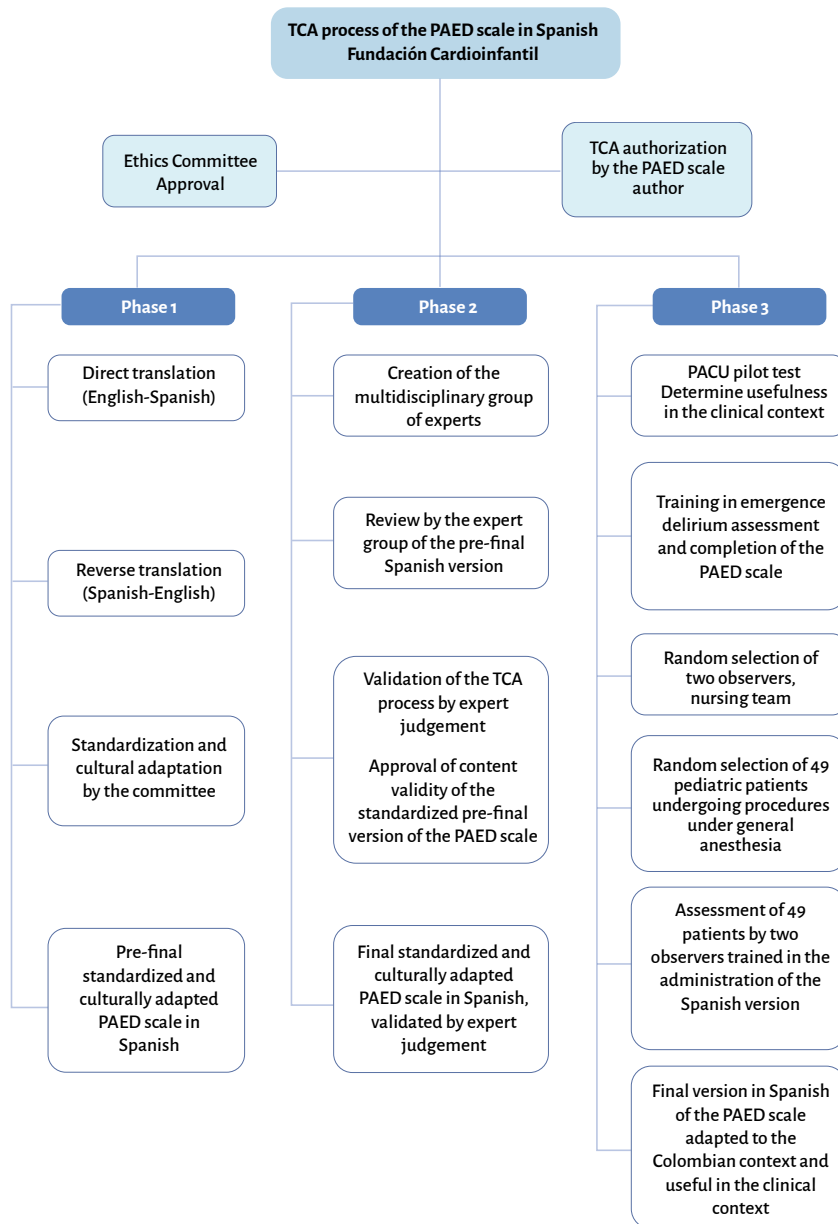
METHODS

Longitudinal prospective study of mixed methodology in which a quantitative and qualitative design was used to approach the research problem, using a descriptive inferential approach. The work was based on a sequential systematic process divided into three phases, each of which consisted of an orderly series of steps (Figure 1). The quantitative component was used for the judgement by experts and the inter-observer agreement analysis, as described

below in the statistical analysis section. The qualitative component was used only during the semistructured interview with the observers, upon completion of the pilot test. This study was approved by the institutional Ethics Committee of Fundación Cardioinfantil-Instituto de Cardiología in May, 2022, as stated in Meeting Minutes No. 027-2022.

Before starting the study, request for authorization to undertake the translation into Spanish of the PAED scale was submitted to Nancy Sikich (7), author of the original scale. This authorization was received on December 6, 2021.

Phases 1 and 2 of the study were carried out outside the institution between April and June, 2022, and phase 3 was carried out in the Pediatric Post-anesthesia Care Unit (PACU) of Fundación Cardioinfantil-IC between May and August, 2023. The different phases of the study were designed in accordance with the unified recommendations of multiple authors and consensuses found in the literature (10-19).

Figure 1. TCA process of the PAED scale in Spanish.

PAED: Pediatric Anesthesia Emergence Delirium, TCA: Translation and cultural adaptation, PACU: Pediatric post-anesthesia care unit.

Source: Authors.

Methodological approach to the translation and cultural adaptation of the PAED scale in Spanish

Phase 1. A descriptive approach was used during this phase, the aim being to maintain the scale's original structure as well as the semantic and experiential equivalence.

Direct translation of the PAED scale from English into Spanish

A certified translation company based in Bogota (Tvains) was commissioned for this purpose. The translation process involved two bilingual translators with Spanish as their native language. The methodology

required the two translators to work independently and blinded to the aims of the study (10-12). One technical and one conventional translation were carried out.

Reverse translation

The translations into Spanish were again sent to the same company for reverse translation, which should be carried out by two bilingual translators whose native language was Canadian English. For this step, a conventional translation was required of the initial technical translation, and a technical translation was required of the initial conventional translation.

Scale unification by the multidisciplinary committee

Once the translations were received, a scale unification committee was set up, integrated by two bilingual experts — anesthetists with experience caring for pediatric patients; a linguist expert in Spanish; a methodologist — epidemiologist — and a member of the research team. The committee met online and used an assessment form to synthesize the translations and identify any important linguistic expressions as compared to the original English version, and finally made the cultural adaptation of the Spanish version. The committee assessed the translations for semantic and experiential equivalence to obtain a unified pre-final, culturally adapted, Spanish version.

Validation by expert judgement of the unified pre-final version of the PAED scale in Spanish

Phase 2 consisted of the review and validation of the pre-final scale TCA process: 1) Setting up the multidisciplinary group of twelve experts, in accordance with the recommendations by Escobar et al. (13) as methodological reference. 2) Review of the unified Spanish version by the expert

group. Each of the assessment categories were reviewed (e.g., restless child) as well as each rating item (e.g., several times) proposed in the new version of the scale. The experts reviewed compliance with four assessment criteria included in the tool designed by Escobar et al. for expert judgement (13), namely: sufficiency (the items in a single domain are sufficient to measure the domain), consistency (specific items are related to the domain they measure), relevance (the items or categories are essential and must be included in the proposed version), and clarity (the included categories are comprehensible and have the adequate semantic and syntactic structure). Additionally, in this form, the experts also determined whether the version of the proposed scale met the content validity criteria. After the one-month time period assigned for the review, each of the members forwarded a completed form to the member of the research team appointed to collect the results. Once all the completed forms were received, the information was processed, a description of the process carried out was prepared, and expert judgement agreement was determined using the quantitative approach.

Pilot test in the PACU of the PAED scale in Spanish

Phase 3. Exploratory pilot test using a descriptive and inferential approach developed across a sequence of steps described below.

Observer selection and training

The majority of the members of the PACU nursing team were trained by means of a brochure and a previously designed explanatory video which included important aspects of ED such as definition, signs, symptoms, clinical assessment considerations, treatment, and preventive measures. This was followed by a random

selection of two members of the nursing team as observers of the pilot test who participated initially in a mock run of the scale and then a practical exercise with real pediatric patients in which they were required to observe for signs and symptoms as well as the diagnosis of ED. After completing the exercise, the observers completed a questionnaire regarding the applicability of the scale, identifying potential barriers to understanding the scale, the instructions, the items, etc.

Pilot test methodological design

The test was performed to determine the usefulness of this scale when applied in a real clinical context. A random patient selection method was chosen in order to improve the external validity of the results, with consecutive case selection based on the institution's scheduled procedures. Additionally, a sample size was calculated using a probabilistic model, with 0.05 confidence and 80% statistical power. The expectation was to achieve moderate to near-perfect inter-observer agreement (k between 0.7 K1 and 0.9 Ko), obtaining a sample size of 45 patients. The pilot test included pediatric patients between 4 and 10 years of age, ASA I, II and III, scheduled for elective procedures under general anesthesia, who understood and could communicate in Spanish. An informed consent form was completed for all patients by the guardian and/or legal representative before their inclusion in the study. Patients with a history of cardiovascular disease and compromised ventricular function or congenital heart disease were excluded, as were also patients with altered behavior and high levels of anxiety during the preoperative period, with or without a prior neurological condition, especially associated with symptoms of restlessness and anesthetic premedication. The above because abnormal behaviors in these patients could be attributed to the underlying condition and not to ED, and could eventually give rise to a selection bias in the pilot test.

Likewise, steps were taken to control other potential biases: random selection of the observers based on voluntary participation to minimize selection; 10 to 15 minute intervals between assessments to avoid information bias, in keeping with the international recommendations; removal of the PAED score cut-off points in the collection tools in order to prevent detection bias, with the score being automatically calculated based on the data entered by the observers; location of observers in separate spaces under the supervision of the research team to avoid communication between them; and standardized training regarding the way to complete, score and interpret the tool in order to ensure homogenous application of the scale. Inter-observer agreement for ED assessment and diagnosis using the Spanish version of the PAED scale was determined by means of systematic observation and assessment of the patients by both observers at four points in time during the first 30 minutes of anesthesia emergence, as follows: T0, immediately upon waking; T1, 10 minutes after waking; T2, 20 minutes after waking; T3, 30 minutes after waking. Using this methodological design in a consecutive random fashion, 49 pediatric patients receiving care in the PACU between May and August 2023 were assessed, thereby achieving complete sample saturation, leading into the next stage of information processing and analysis.

Semistructured interview

To avoid memory bias, the qualitative methodology was applied immediately after completion of the pilot test to conduct a semi-structured interview of each the observers who participated in the pilot test. The aim of the interview was to learn about their experience with the study and gather information regarding the usefulness and acceptability of the Spanish version of the PAED scale. The answers provided by the respondents were recorded and then transcribed verbatim. A thematic analysis

of the transcription was then conducted to identify the information pertaining to the themes and categories studied by the interview. Once these were identified, a deductive interpretation of the results was made and the conclusions regarding the collected information were prepared. This information is presented in narrative form in the results section. The following are the situations and categories subject to analysis as part of the semi-structured interview: 1) the experience of the parties involved during the participation in the study was analyzed according to the following situations: challenges faced during the use of the scale and the particular experience of each of the reviewers. 2) The usefulness of the scale was analyzed in accordance with the following categories: time to complete the tool, assessment timing, clinical relevance of the proposed version, acceptability in the clinical setting, items that created confusion or were difficult to interpret, and final suggested changes to the scale.

Information gathering tools

For this pilot test, an electronic form was designed in the database of the institution's anesthesia service using the Claris Filemaker Pro-19 software, which permitted data gathering from the observers' mobile devices, controlling for answer replication bias, given that each record leaves a unique user digital footprint. The results were exported to Excel (Microsoft Office 2019), and the database was reviewed, debugged, re-coded and later exported to the SPSS IBM version 26 software package for statistical analysis.

Statistical analysis

The statistical analysis took into consideration the fact that the PAED is a Likert-type scale consisting of 5 assessment categories and 5 measurement items of ordinal type, with numerical values

(0 to 4) organized in descending order for assessment categories 1 to 3 and in ascending order for categories 4 and 5 (7). This particular characteristic of the structure was not modified during the process of analyzing and interpreting the information.

Expert judgement agreement was analyzed using Fleiss' k index, and inter-observer agreement during the pilot test was analyzed using Cohen's k index. A confidence of 95% was selected for the confidence intervals (CI). The p value was considered statistically significant when less than 0.05. The interpretation of the degree of agreement according to the k index value obtained was done in accordance with the Landis and Koch criteria (17-19).

Inter-observer agreement for the diagnosis of ED using the proposed scale was analyzed using the intraclass correlation coefficient (ICCA), which considers any difference across observations as disagreement: As inter-observer disagreement increases, the ICCA value decreases accordingly. In this case, the decision was to assess absolute agreement for each of the results on the scale at every point in time, selecting the two-factor random effects analysis model to this end.

The conventional values used to interpret the ICCA were (20,21): <0.40, poor; 0.40-0.59, sufficient; 0.60-0.74, good, and 0.75-1, excellent.

In order to establish a statistical precedent regarding ED detection using the Spanish version of the PAED scale assessed in this study, the proportion of patients with a score of 10 or more on the scale was calculated for exploratory purposes. The calculation was made after completing the test, with the aim of identifying the patients with a condition consisting with ED at the predetermined measurement time points (7).

To this end, the patients who obtained said score, as well as the observer and the corresponding assessment time points were identified. This number was divided

by the total number of patients in order to derive the proportion. It is worth noting that the pilot test was designed to evaluate the observers as they used the Spanish version of the proposed scale during patient assessments; the priority was not to detect whether treatment of an ED episode was triggered by the assessment or administered as part of an immediate care protocol. The total score obtained did not influence medical action after the assessment was conducted.

RESULTS

Unification of the Spanish version of the PAED scale

After receiving the translations, the committee met to unify the scale and make the cultural adaptation. The committee decided not to adopt the translated versions literally but to introduce the following semantic modifications for the sake of clarity, improved flow and efficient use of the Spanish language: 1) a distinction was made between boys and girls for each assessment category; 2) because of semantic differences between translations in the second category, the committee decided to call it "purposeful actions"; 3) bearing in mind that the subjects of assessment will find themselves in an unknown surrounding, the structure of the third category was changed to read "the boy/girl recognizes his/her surroundings", as it implies a greater degree of alertness.

After introducing the changes suggested by the unification committee, a unified pre-final version of the PAED scale was created (Table 1), which was then submitted to expert judgement at a later stage.

Expert judgement validation of the pre-final Spanish version of the PAED scale

The expert group was multidisciplinary, in accordance with the recommendations by

Escobar et al. (13). This was reflected in the diversity of professionals and specialists and their years of experience, ranging between 4 and 30, with a median of 10 years (supplementary material 1).

Expert judgement found that the pre-final assessment categories met the sufficiency, relevance and clarity criteria and were, therefore, validated by expert judgement, with considerable to near-perfect agreement (Table 2). Likewise, the expert group found, with a moderate to near-perfect agreement, that the measurement items proposed in the version of the scale under assessment, met the sufficiency, consistency, relevance and clarity criteria (Table 3), approving and validating these items as measurements of each of the assessment categories in the scale.

Thus were the proposed objectives for this phase of the study accomplished, resulting in the final Spanish version of the PAED scale and the approval of both the TCA process as well as content validity.

Analysis of inter-observer agreement during the pilot test

This pilot test was exploratory in nature. An inter-observer agreement analysis was conducted as the two observers were using the final Spanish version of the PAED scale to assess the 49 patients included in the study (Table 4) at the predetermined time points during emergence from anesthesia.

To: A high degree of inter-observer agreement was achieved for assessment categories 1, 2, 4 and 5. However, the degree of agreement dropped to acceptable for category 3, indicating greater variability in the observations. With high agreement, the observers found that most patients made eye contact on multiple occasions, their actions were purposeful several times and they were never found to be restless or inconsolable. With acceptable agreement, it was found that they recognized their surroundings several times.

T1: High inter-observer agreement was obtained, and the same trend as for To was

Table 1. Unified version of the PAED scale.

Category	Assessment item description	Measurement items					Total score
		Never	Few times	Several times	Many times	Always	
1	The boy/girl makes eye contact with the caregiver	4	3	2	1	0	
2	The boy/girl's actions are purposeful	4	3	2	1	0	
3	The boy/girl recognizes his/her surroundings	4	3	2	1	0	
4	The boy/girl is restless	0	1	2	3	4	
5	The boy/girl is inconsolable	0	1	2	3	4	
Total							

Source: Authors, derived from the Scale Unification Committee.

Table 2. Results of the category assessments with the adapted PAED scale.

	Fleiss' k	p value	Confidence interval	
Sufficiency	0.765	0.001	0.72	0.81
Consistency	0.873	0.001	0.83	0.92
Relevance	0.794	0.001	0.75	0.84
Clarity	0.661	0.001	0.62	0.71

Source: Authors.

Table 3. Assessment results for the measurement items of the adapted PAED scale.

	Fleiss' k	p value	Confidence interval	
Sufficiency	0.618	0.001	0.58	0.66
Consistency	0.589	0.001	0.55	0.63
Relevance	0.742	0.001	0.70	0.79
Clarity	0.535	0.001	0.49	0.58

Source: Authors.

noted, with a noticeable drop in agreement for category 3 assessments.

T2: High inter-observer agreement was achieved in all the assessment categories of the scale at this time point during emergence. Most of the patients made eye contact with the caregiver many times, their actions were purposeful many times, and they were never found to be agitated or inconsolable.

T3: Substantial to near-perfect agreement was obtained in the different assessment categories at this time point. As a result, 30 minutes after anesthesia emergence, the observers found that most patients made eye contact with the

Table 4. Characteristics of the patients observed during the pilot test.

Characteristic	n (%)
Age, months (SD)	6.4 (1.6)
Sex	
Female	23 (46.9)
Male	26 (53)
ASA	
I	6 (12.2)
II	16 (32.6)
III	27 (55)
Procedure type	
Hematology-Oncology	29 (59.1)
Urology	13 (26.5)
Pediatric surgery	7 (14.2)

SD: Standard deviation.

Source: Authors

Table 5. Absolute agreement for the total score on the PAED scale.

	ICC	Confidence interval	
PAED scale result T0	0.92	0.846	0.96
PAED scale result T1	0.88	0.785	0.935
PAED scale result T2	0.82	0.693	0.902
PAED scale result T3	0.87	0.760	0.931

ICC: Intra-class correlation coefficient.

Source: Authors.

caregiver all the time, their actions were always purposeful, they recognized their surroundings at all times and they were never found to be agitated or inconsolable.

Next, inter-observer agreement during the pilot test for the diagnosis of ED was evaluated according to the scale cutoff points consistent with this diagnosis. Agreement was found to be considerable to near-perfect, with Cohen's k ranging between 0.77 and 1 (supplementary material 2). For T0 there were 9 agreements and 2 disagreements between the observers, for T1 there were 4 agreements and 2 disagreements, while there were 0 disagreements for the last two time points.

Additionally, the total score obtained by each of the observers using the scale was also analyzed, with a finding of absolute agreement between the two.

Absolute agreement was excellent, with ICCA values ranging between 0.82 and 0.92. The results are presented in Table 5. Although this was not the main objective of the pilot test, the sample size allowed to determine excellent inter-observer agreement, indicating that diagnosis was reliable, with low disagreement.

The analysis carried out a posteriori of the proportion of patients identified as having a diagnosis consistent with ED, showed that this proportion ranged between 16% for observer 1 and 18.3% for observer 2, when patients were assessed immediately upon waking from anesthesia; this was the highest recorded proportion

(supplementary material 3). The proportion of patients in whom a case consistent with ED was identified dropped rapidly over the first 30 minutes and dropped by more than 50% after 10 minutes (9%), and it was found to stabilize at 2% 30 minutes into the postoperative period.

Semi-structured interview with each of the observers

Following the completion of the pilot test, the semi-structured interview was used to gather feedback from the observers regarding the use of the scale during the pilot test. The results were as follows: 1) The observers agreed that the proposed version of the PAED scale is useful in the PACU; 2) completion of the scale takes less than 2 minutes, making it suitable for the pace and needs of patients receiving care at the Fundación Cardioinfantil-IC PACU; 3) the scale was comprehensible and easy to complete; 4) the observers found that the predetermined time points at which to assess the patients over the first 30 minutes of emergence were adequate, allowing for timely identification of cases consistent with ED.

DISCUSSION

The PAED scale has been translated and adapted in several languages, including

Danish, Portuguese and Spanish. The existing Spanish adapted version was developed with a specific cultural adaptation for Chile (22). Bearing in mind the evidence found in the literature regarding differences in cultural characteristics among countries, regions and cities where the same language is spoken, which introduce variations into the psychometric properties of a tool (10-14), we considered it advisable to abide by such recommendation and undertake the process of translation and cultural adaptation, thus ensuring the relevance and validity of the PAED scale in relation to the Colombian dialect.

The analysis of the study by Epulef et al. (22) showed that these authors took the literal translation of the scale without introducing any linguistic expressions either in the categories or the items. For this study, the linguists who participated in the unification committee suggested linguistic expressions and semantic changes to improve clarity, flow and efficiency in the use of the Spanish language in the Colombian context.

Categories 1 and 5 of the assessment in the translation by Epulef et al. (Chilean version) were kept, given that no major semantic differences were identified; the flow of language in the other three categories is distinctly Colombian Spanish. The repercussions of these modifications were evident and confirmed in the results of the semi-structured interview in which the observers did not suggest any changes to the scale, as they found that the assessment categories and measurement items were clear.

The methodological strengths of this study include the scale unification process carried out by a multidisciplinary team, taking into account various published international consensus recommendations (10-19), thus allowing for adequate adaptation to the Colombian linguistic context. Moreover, the pre-final version of the PAED scale was submitted to expert judgement to determine content validity and to validate the process of translation

and cultural adaptation (TCA) in Spanish. With moderate to near-perfect agreement, the experts approved the previous process judging that the version obtained met the sufficiency, consistency, relevance and clarity criteria to a high degree (13). All of the methodological aspects described are synergistic, adding scientific rigor and external validity to the work.

Barbosa et al. developed a similar TCA process with the CAPD scale with the aim of identifying delirium in any type of pediatric population in intensive care. Methodological differences were identified when comparing it with the TCA process carried out for the PAED scale: they created a panel of 16 experts which acted as unification committee and produced a pre-final version for study, without estimating degrees of agreement among the experts during the assessment process (23).

In this study, we chose a quantitative and inferential approach designed to obtain results which could be eventually replicated by other authors, and also to create evidence for future studies aimed at validating the clinical replicability of the Spanish version of the PAED scale. It is worth noting that such an approach was not used in the publications which deal with translation and transcultural adaptation processes for the PAED scale currently available in the literature (22,24,25).

Later, during phase 3 of the study, a pilot test using the final Spanish version of the scale was conducted. It is worth noting that the aim of this phase was not to compare the observation against a gold standard, but to determine how close were the observations to a diagnosis consistent with ED when the observers used the Spanish version at the four predetermined time points in 49 pediatric patients receiving care in the PACU.

It is notable that inter-rater agreement improved progressively across the five proposed assessment categories increased over the course of the assessment time period. Interestingly, the agreement results obtained for the last assessment category were similar to those described by

Ringblom et al. (26) who reported a k index of 0.91 (95% CI: 0.86-0.95).

The absolute inter-observer agreement for the diagnosis of ED based on the total score for each patient was excellent. These results are consistent with the previously reported agreement for the five assessment criteria and are similar to the reports found in the literature. In the study by Sikich et al. (7), inter-observer reliability using the original version of the PAED scale was 0.84 (95% CI: 0.76-0.90), showing excellent agreement across observations. In the version by Epulef et al., overall agreement among the reviewers was 0.69 (95% CI: 0.60-0.77) and 0.66 (95% CI: 0.30-1.00) among the anesthetists, falling in the good agreement group (22).

Also, in line with the results of the semi-structured interview, the pilot test allowed to assess acceptance of the scale by the staff in the unit, as well as usefulness in terms of form simplicity, application team, ease of completion, comfortable use, and high relevance for the construct to be measured. Therefore, acceptability and usefulness of this version was confirmed during the first formal application of the scale in the clinical context.

The process described in previous sections based on the estimated sample size showed promising results, as evidenced by the high degree of inter-observer agreement and low variability. It goes without saying that the adequate use and interpretation of the scale also helped in some way to reduce the inherent variability of the measurement tool. This underscores the need to use a larger sample size for an in-depth study of the Spanish version of the PAED scale in order to gain accurate insight into its variability as a diagnostic tool and its actual contribution when it comes to diagnosing ED in the Colombian context.

Though not the primary measurement in this study, the proportion of patients identified as cases consistent with ED — derived from the pilot test — behaved similarly as in other studies, albeit in different languages. Simonsen et al. (24) reported a prevalence of ED of 13.2% in

his study of the PAED scale in Danish, and Quintão et al. (25) 18% with the Portuguese version. Those results are consistent with ours, despite potential differences in the populations or methodologies used in those studies. A score of 10 or more on the PAED scale was used for the official diagnosis of ED in all cases (7).

It is imperative to recognize various biases which could have influenced the quality of the responses and the interpretation of the results. The observers involved in applying the tool did so during their workday in the PACU, subject to the various distractors inherent to the clinical environment, which could have interfered with their ability to pay close attention to the responses as they assessed the patients.

Additionally, after completion of the pilot test and during the semi-structured interview, it was found that the observers had difficulty understanding cutoff points for the diagnosis of ED. Despite this lack of clarity, results were strong, with high inter-observer agreement. It is worth noting that this uncertainty originated in the way the observers adopted the cutoff points. Therefore, it is recommended to reinforce cutoff values for the diagnosis of ED during staff training. This recommendation should be considered for future research designed to strengthen the validity of the results derived from this test.

CONCLUSION

This translation and cultural adaptation process met the proposed objectives, leading to the development of a final version of the PAED scale in Spanish which can be used in the Colombian clinical context as a tool for the timely diagnosis of ED in the pediatric population. This work is a sound first step in the proposed approach and establishes a strong initial basis for future research focusing on strengthening the validity and replicability of the new version of the PAED scale in the pediatric population in Colombia, thus supporting its usefulness in the clinical setting. This version was shown to be useful, practical and valuable for assessing emergence

delirium, with excellent inter-observer agreement in the pilot test.

ETHICAL CONSIDERATIONS

Endorsement from the ethics committee

This study was approved by the ethics committee of Fundación Cardioinfantil-Instituto de Cardiología in May 2022, as stated in Minutes 027-2022.

Protection of human and animal subjects

The authors declare that no experiments were performed on humans or animals for this study. The authors declare that the procedures followed were in accordance with the ethical standards of the responsible human experimentation committee and with those of the World Medical Association and the Declaration of Helsinki.

Data confidentiality

The authors declare that they have followed the protocols of their work center on the publication of patient data.

Right to privacy and informed consent

The authors declare that no patient data appear in this article.

The authors have obtained the written informed consent of the patients or subjects mentioned in the article. The corresponding author is in possession of said document.

ACKNOWLEDGEMENTS

Authors' contributions

All the authors contributed to the study conception and design, development of

the research protocol, data analysis and interpretation, and writing and critical review of the manuscript.

Advice for the development of the study

None declared.

Funding

None declared.

Conflict of interest

None declared.

Presentations

None declared.

Appreciation

We are grateful to Jorge Humberto Mayorga-Álvarez, statistician, Master's in Statistical Science (expert in scale validations), Universidad Nacional de Colombia and Álvaro Eduardo Osorio, pediatric psychiatrist, Universidad Nacional de Colombia, Pontificia Universidad Javeriana.

REFERENCES

1. Abdelaziz H, Bakr R, Kasem A. Effect of intranasal dexmedetomidine or intranasal midazolam on prevention of emergence agitation in pediatric strabismus surgery: a randomized controlled study. *Egyptian J Anaesth*. 2016;32(3):285-91. <https://doi.org/10.1016/j.egja.2015.11.009>.
2. Yuen V. Dexmedetomidine: perioperative applications in children. *Pediatric Anesthesia*. 2010;20(3):256-64. <https://doi.org/10.1111/j.1460-9592.2009.03207.x>.
3. Sampson EL, West E, Fischer T. Pain, and delirium: mechanisms, assessment, and management. *Eur Geriatr Med*. 2020;11(1):45-52. <https://doi.org/10.1007/s41999-019-00281-2>.

4. Dahmani S, Delivet H, Hilly J. Emergence delirium in children. *Curr Opin Anaesthesiol*. 2014;27(3):309-15. <https://doi.org/10.1097/aco.000000000000076>.
5. Vercueil A. Ischaemic conditioning: Intervening to protect; Before, after, and at a distance. *Anesthesia*. 2015;70(4):379-83. <https://doi.org/10.1111/anae.13054>.
6. Sethi S, Ghai B, Ram J, Jd W. Postoperative emergence delirium in pediatric patients undergoing cataract surgery – a comparison of desflurane and sevoflurane. *Pediatric Anesthesia*. 2013;23(12):1131-7. <https://doi.org/10.1111/pan.12260>.
7. Sikich N, Lerman J. Development and psychometric evaluation of the pediatric anesthesia emergence delirium scale. *Anesthesiology*. 2004;100(5):1138-45. <https://doi.org/10.1097/00000542-200405000-00015>.
8. Sánchez MPG, Franco-Gruntorad DG. Frecuencia de delirium de emergencia en niños llevados a procedimientos hemato-oncológicos bajo anestesia general, Bogotá 2019 [tesis de especialización]. [Bogotá]: Universidad del Rosario; 2019. https://doi.org/10.48713/10336_20915.
9. Banchs RJ, Lerman J. Preoperative anxiety management, emergence delirium, and postoperative behavior. *Anesthesiol Clin*. 2014;32(1):1-23. <http://dx.doi.org/10.1016/j.anclin.2013.10.011>.
10. Ramada-Rodilla JM, Serra-Pujadas C, Delclós-Clanchet GL. Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas. *Salud Publica Mex*. 2013;55(1):57-66. <http://dx.doi.org/10.1590/S0036-36342013000100009>.
11. Luján-Tangarife JA, Cardona-Arias JA. Construction and validation of measurement scales in health: a review of psychometric properties. *iMedPub Jour*. 2015;1-10. <https://dialnet.unirioja.es/servlet/articulo?codigo=5178935>.
12. Soriano Rodríguez AM. Design and validation of measurement instruments. *Diálogos*. 2015;(14):19-40. <http://dx.doi.org/10.5377/dialogos.v0i14.2202>.
13. Escobar-Pérez J, Cuervo-Martínez Á. Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición*. 2008(6); 27-36.

14. Cortés-Reyes É, Rubio-Romero JA, Gaitán-Duarte H. Statistical methods for evaluating diagnostic test agreement and reproducibility. *Rev Colomb Obstet Ginecol.* 2010;61(3):247-55. <http://dx.doi.org/10.18597/rcog.271>.
15. Sánchez R, Echeverry J. Validación de escalas de medición en salud. *Rev Salud Pública (Bogotá).* 2004;6(3):302-18. <http://dx.doi.org/10.1590/s0124-00642004000300006>.
16. Torres-Gordillo JJ, Perera-Rodríguez VH. Cálculo de la fiabilidad y concordancia entre codificadores de un sistema de categorías para el estudio del foro online en e-learning. *Rev Investigación Educat.* 2009;27(1):89-103. <https://www.redalyc.org/articulo.oa?id=283322804006>.
17. Carvajal A, Centeno C, Watson R, Martínez M, Sanz Rubiales Á. How is an instrument for measuring health to be validated?. *An Sist Sanit Navar.* 2011;34(1). <http://dx.doi.org/10.4321/s1137-66272011000100007>.
18. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quinónez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Front Public Health.* 2018;6. <http://dx.doi.org/10.3389/fpubh.2018.00149>.
19. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60(1):34-42. <http://dx.doi.org/10.1016/j.jclinepi.2006.03.012>.
20. Martínez Pérez JA, Pérez Martin PS. Coeficiente de correlación intraclase. *Semergen.* 2023;49(3):101907. <http://dx.doi.org/10.1016/j.semerg.2022.101907>.
21. Método UAB: Estudios de Formación Continuada. Coeficiente de correlación intraclase [Internet]. Universidad Autónoma de Barcelona. https://metodo.uab.cat/arxiu/material/general/El_coeficiente_de_correlacion_intraclase.pdf
22. Epulef V, Muñoz S, Alarcón AM, et al. Spanish version of the Pediatric Anesthesia Emergence Delirium scale: translation and cross-cultural adaptation. *BMC Anesthesiol.* 2022;22(1). <https://doi.org/10.1186/s12871-022-01893-1>.
23. Barbosa M, Duarte M, Bastos V, Andrade L. Translation and cross-cultural adaptation of the Cornell assessment of pediatric delirium scale for the Portuguese language. *Rev Bras Ter Intensiva.* 2018;30(2). <http://dx.doi.org/10.5935/0103-507x.20180033>.
24. Simonsen BY, Skovby P, Lisby M. An evaluation of the Danish version of the Pediatric Anesthesia Emergence Delirium scale. *Acta Anaesthesiol Scand.* 2020;64(5):613-9. <http://dx.doi.org/10.1111/aas.13543>.
25. Quintão VC, César R, Carmona M, Ferreira A. A psychometric evaluation of the Brazilian versions of the Pediatric Anesthesia Emergence Delirium scale and Children's Hospital of Eastern Ontario Pain Scale. *Paediatr Anaesth.* 2021;31(12):1366-7. <http://dx.doi.org/10.1111/pan.14300>.
26. Ringblom J, Wählin I, Proczkowska M. A psychometric evaluation of the Pediatric Anesthesia Emergence Delirium scale. *Paediatr Anaesth.* 2018;28(4):332-7. <http://dx.doi.org/10.1111/pan.13348>.

SUPPLEMENTARY MATERIAL

Supplementary material 1. Expert group profile.

Expertise	Mean	Median
Professional experience in years	12±8.1	10.5
Range in years	Min 4	Max 30
Professions of the experts in the group	Frequency	%
Cardiovascular anesthetist	4	33.3
Pediatric anesthetist	3	25.0
Anesthetist-Methodologist	1	8.3
Registered Nurse	1	8.3
Bachelor in Spanish and Literature	1	8.3
Psychologist-Methodologist	1	8.3
Pediatric psychiatrist	1	8.3

Source: Authors.

Supplementary material 2. Agreement in the diagnosis of ED.

	Agreements Freq. (%)	Disagreement Freq. (%)	Cohen's k	p
To	9 (18.3)	2 (4)	0.875	0.000
T1	4 (8)	2 (4)	0.778	0.000
T2	1 (2)	0	1	0.000
T3	1 (2)	0	1	0.000

Freq.: Absolute frequency.

Source: Authors.

Supplementary material 3. Proportion of patients with a diagnosis consistent with ED.

	Observer 1 Freq. (%)	Observer 2 Freq. (%)
To	9 (18.3)	8 (16)
T1	4 (8)	6 (12)
T2	1 (2)	1 (2)
T3	1 (2)	1 (2)

Freq.: absolute frequency.

Source: Authors.