

Study Protocol and Statistical Analysis Plan

Inter-Rater Agreement of Four Behavioral Rating Scales in Children Undergoing Dental Treatment Based on Clinical Video Recordings

Official title: Inter-Rater Agreement of Four Behavioral Rating Scales in Children Undergoing Dental Treatment Based on Clinical Video Recordings

Short title: Inter-Rater Agreement of Pediatric Dental Behavioral Scales

ClinicalTrials.gov identifier: To be assigned

Protocol/SAP version: 1.0

Document date: 04 September 2026

Sponsor: Universitat de València

Study type: Observational

Study model: Prospective, single-center, descriptive observational study

Time perspective: Prospective enrollment with retrospective video assessment

Part I. Study Protocol

1. Study synopsis

This single-center prospective observational study will evaluate inter-rater agreement among four behavioral rating scales used in pediatric dentistry: the Houpt Scale, the Ohio State University Behavioral Rating Scale (OSUBRS), the Venham Scale, and the Frankl Scale.

Approximately 30 children aged 4–10 years who require clinically indicated restorative treatment for dental caries under local anesthesia will be recruited consecutively at the University Dental Clinic of Universitat de València. Selected stages of routine dental treatment will be video-recorded after written parental or legal-guardian consent. Three trained evaluators who will not participate in the dental treatment or recording will independently assess the recordings.

The clinical treatment will not be assigned, modified, delayed, or withheld for research purposes. The research procedures will consist of video and audio recording of predefined stages of routine care and subsequent behavioral assessment.

2. Background and rationale

Reliable behavior assessment can help dentists make clinical decisions, choose suitable behavior-management techniques, and assess whether these techniques improve children's cooperation during dental treatment.

Several behavioral rating scales are used in pediatric dentistry, including the Houpt Scale, OSUBRS, Venham Scale, and Frankl Scale. These instruments differ in their structure, level of detail, and practical requirements. Their usefulness in clinical practice and research depends partly on reproducibility and agreement between independent evaluators.

Video recordings of real clinical situations allow several evaluators to assess the same standardized events retrospectively. This approach permits comparison of the inter-rater agreement and practical performance of four commonly used behavioral rating scales under comparable clinical conditions.

3. Objectives

3.1 Primary objective

To evaluate inter-rater agreement among three trained evaluators applying the Houpt Scale, OSUBRS, Venham Scale, and Frankl Scale to video recordings of children aged 4–10 years undergoing restorative dental treatment under local anesthesia.

3.2 Secondary objectives

To explore whether agreement varies according to age, sex, treated tooth, or previous dental treatment experience, where the available sample permits meaningful analysis.

4. Study design

This is a prospective, single-center, descriptive observational study involving prospective enrollment and video recording followed by retrospective behavioral assessment.

Participants will receive restorative dental treatment for caries as part of routine clinical care. The research team will not determine the clinically indicated treatment and will not alter the treatment plan. Participants will be enrolled consecutively after ethics approval and after written informed consent has been obtained from a parent or legal guardian.

The study will use a single observational cohort. There will be no randomization, control group, study arm, placebo, investigational product, or research-assigned clinical intervention.

5. Study setting

The study will be conducted at the University Dental Clinic of the Fundació Lluís Alcanyís, Universitat de València, Department of Stomatology, Gascó Oliag No. 1, Valencia, Spain. Clinical procedures will take place in the postgraduate pediatric dentistry clinic.

6. Study population

The study population will comprise children aged 4–10 years who attend the clinic for clinically indicated restorative treatment of dental caries requiring local anesthesia.

6.1 Inclusion criteria

Participants must meet all of the following criteria:

- Aged 4–10 years, inclusive.
- American Society of Anesthesiologists physical status I or II.
- Requiring restorative treatment for dental caries under local anesthesia.
- Attending the participating university dental clinic.

6.2 Exclusion criteria

Participants will be excluded if any of the following apply:

- Neurological or cognitive alteration that could affect the planned behavioral assessment.
- American Society of Anesthesiologists physical status III or IV.
- Failure to obtain any required parental or legal-guardian consent.
- Clinical circumstances in which recording could interfere with patient safety, treatment, or well-being.
- Video recording of insufficient quality to permit behavioral assessment.
- Withdrawal of consent before completion of study procedures.

7. Recruitment and informed consent

Potentially eligible children will be identified among patients attending the clinic for examination or dental treatment. Parents or legal guardians of children who meet the eligibility criteria will be informed about the purpose of the study, study procedures, voluntary participation, video and audio recording, data-protection measures, potential risks, and the right to withdraw without consequences for the child's treatment.

Written informed consent will be obtained before recording begins. Refusal to participate will not affect the dental care received. Child assent will be obtained when appropriate for the child's age and level of understanding.

8. Clinical treatment and recording procedures

The dental treatment will be provided by postgraduate pediatric dentistry students who are qualified and licensed dentists, under the supervision required by the clinic. The treatment will consist of restorative management of dental caries requiring local anesthesia and will follow routine clinical practice.

Video recording will be performed only after the required consent has been obtained. Recordings may be obtained using a ceiling- or wall-mounted camera owned by the Fundació Lluís Alcanyís-UV. Recordings may include image and sound to permit assessment of verbal and non-verbal responses and the interaction between the child and dental professional.

Recording may be activated by the treating professional or assistant. Recording will be stopped, omitted, or not initiated whenever the treating professional considers that it could interfere with patient safety, treatment, or well-being.

The following four treatment stages will be assessed:

1. Reclining the dental chair.
2. Administration of infiltrative local anesthesia.
3. Placement of the rubber dam.
4. Use of rotary dental instruments.

Behavior will be assessed during a two-minute observation period for each stage, when that stage is available and sufficiently captured in the recording.

The recordings are solely for research purposes. They will not be used for teaching, public display, or publication of identifiable images or audio.

9. Behavioral rating scales

Three evaluators who will not participate in the child's treatment or recording will independently assess the video recordings. Evaluators will be blinded to one another's ratings and will use only the study code and relevant recording information.

9.1 Houpt Scale

The following Houpt domains will be recorded:

1. Consciousness: fully awake and alert; drowsy or disoriented; asleep.
2. Movement: violent and interrupts treatment; continuous and makes treatment difficult; controllable and does not interfere with treatment; no movement.
3. Crying: hysterical and demanding attention; continuous and persistent, making treatment difficult; intermittent and mild, not interfering with treatment; no crying.
4. Overall behavior: treatment abandoned; poor and only partially completed; fair and completed after interruption; good but difficult; very good with limited crying or movement; excellent, with no crying or movement.

9.2 Ohio State University Behavioral Rating Scale

1. Quiet.
2. Crying without movement.
3. Movement without crying.
4. Struggling or active resistance.

9.3 Venham Scale

0. Full cooperation.
1. Mild protest or soft crying indicating discomfort without interfering with progress.
2. More prominent protest, crying, hand signals, or head movement that may make treatment more difficult.
3. Protest creates a real problem for the dentist.
4. Protest interrupts the procedure and requires the dentist's full attention.
5. Generalized protest, with no compliance or cooperation, requiring physical restraint.

9.4 Frankl Scale

1. definitely negative: Refusal of treatment and intense crying.
2. negative: Reluctance to accept treatment and uncooperative behavior.
3. positive: Acceptance of treatment and generally cooperative responses, sometimes with reservations.
4. definitely positive: Good relationship with the dentist, interest in the procedure, and enjoyment of the situation.

10. Evaluator training and calibration

Before assessing study recordings, evaluators will receive training in the use of all four scales. Training will include discussion of seven videos of children who will not be included in the study.

The three evaluators will then be calibrated by independently assessing five cases and reviewing their ratings with the principal investigator. The calibration procedure will be documented. Agreement during calibration will be assessed before the main assessment phase.

11. Variables and data collection

The following variables will be collected:

- Study identification code.
- Age in years.
- Sex
- Treated tooth.
- Previous dental treatment experience.
- Treatment stage.
- Behavioral score assigned by each evaluator for each scale.

The primary study variables are the behavioral scores assigned independently by the three evaluators for each treatment stage and each scale.

No names or direct personal identifiers will be included in the analytical dataset.

12. Data management and confidentiality

Each participant will be assigned a numerical study code. The document linking the code to the participant's identifying information will be stored separately from the video files and analytical dataset in a secure, access-restricted location. Only the principal investigator, or an authorized person in justified circumstances, may re-identify a participant. Any access to identifying information will be documented.

Video files will be labelled using study codes and sequential identifiers and will not contain participant names. Access to audiovisual material will be restricted to authorized members of the research team. Data will be handled confidentially using appropriate technical and organizational safeguards, including secure storage and encryption where available.

The videos will be used exclusively for the approved research purposes. No identifiable videos, photographs, audio recordings, or individual-level visual data will be shared publicly or with external researchers. Results will be reported only in aggregate form, without identifying participants.

The original video recordings will be permanently deleted five years after the beginning of the research, or earlier if required by the approved data-management plan or applicable institutional policy.

13. Ethical considerations

The study has been approved by the Human Research Ethics Committee of the Universitat de València (2026-ODON-4511724). Recruitment and recording will begin only after the study has been registered on ClinicalTrials.gov.

The study involves no research-assigned change to the dental treatment. The principal additional research procedure is video and audio recording during selected stages of routine dental care. The principal foreseeable risk concerns confidentiality and protection of the child's image and voice. These risks will be mitigated through explicit parental consent, restricted access, pseudonymization, secure storage, and deletion of recordings according to the approved retention period.

There is no expected direct benefit to participants. The expected scientific and societal benefit is improved knowledge of the reliability and comparability of behavioral rating scales used in pediatric dentistry.

14. Sample size justification

The planned sample size is 30 participants. Agreement among three evaluators for ordinal categorical variables will be assessed using Fleiss' kappa and weighted kappa, as appropriate.

The initial sample-size assessment was based on a two-sided alpha of 0.05, 80% power, a null kappa of 0.00, and an alternative kappa of 0.60. The calculation indicated a minimum of 15 participants for the most conservative four-category scenario and 9 participants for the six-category scenario. After allowing for approximately 15% of potentially unusable recordings or withdrawals, the adjusted minimum was 18 participants.

The target sample will be increased to 30 participants to improve the stability and precision of agreement estimates, permit separate analyses of the four Houpt domains, increase the likelihood that a range of behavioral responses will be represented, and provide protection against technical recording problems, clinically necessary interruptions, or other exclusions.

15. Timeline

- 20–30 October 2026: Recording of at least seven non-study training cases.

- 1 -15 November 2026: Evaluator training.
- 15-30 November 2026: Evaluator calibration.
- 1 December 2026- 30 March 2027: Recording of study participants, with a target of 30 children.
- 1 April- 30 May 2027: Evaluation of video recordings.
- 1 June- 23 July 2027: Statistical analysis and study completion activities.

The dates are anticipated and may be updated recruitment or clinical scheduling changes.

16. Dissemination and data sharing

Results may be submitted for publication in scientific journals and presented at scientific meetings. Only aggregate, non-identifiable results will be disseminated.

Individual participant data and video recordings will not be shared with other researchers because the material includes sensitive audiovisual data from minors, including potentially identifiable faces and voices. The approved data-management plan restricts access to the research team and does not authorize external distribution of videos or identifiable individual-level data.

17. References

1. Orellana Calderón et al. Evaluación de la eficacia de las escalas de comportamiento en odontopediatría: revisión de la literatura. *Anatomía Digital*. 2024;7(3.2):118–141.
2. Garret-Bernardin A, Festa P, Matarazzo G, Vinereanu A, Aristei F, Gentile T, et al. Behavioral modifications in children after repeated sedation with nitrous oxide for dental treatment: a retrospective study. *International Journal of Environmental Research and Public Health*. 2023;20(4):4037.
3. Silva JA, Paiva SM, Pereira DMT, Maia LC, Primo LG, Fonseca-Gonçalves A. Are behavior rating scales able to identify behavioral changes in preschool children undergoing a dental intervention? A systematic review. *Pesquisa Brasileira em Odontopediatria e Clínica Integrada*. 2020;20:e5400. doi:10.1590/pboci.2020.097.
4. Sado-Filho J, Viana KA, Corrêa-Faria P, Costa LR, Costa PS. Randomized clinical trial on the efficacy of intranasal or oral ketamine-midazolam combinations compared to oral midazolam for outpatient pediatric sedation. *PLOS ONE*. 2019;14(3):e0213074. doi:10.1371/journal.pone.0213074. PMID: 30856181; PMCID: PMC6411109.
5. Venham LL, Gaulin-Kremer E, Munster E, Bengston-Audia D, Cohan J. Interval rating scales for children's dental anxiety and uncooperative behavior. *ASDC Journal of Dentistry for Children*. 1980;47(1):22–31.

6. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
7. McHugh ML. Interrater reliability: the kappa statistic. *Biochemia Medica*. 2012;22(3):276–282.
8. Bujang MA, Baharum N. Guidelines of the minimum sample size requirements for Kappa agreement test. *Epidemiology, Biostatistics and Public Health*. 2017;14(2):e12267.
9. Donner A, Eliasziw M. A goodness-of-fit approach to inference procedures for the kappa statistic. *Statistics in Medicine*. 1992;11(9):1155–1168.
10. Flack VF, Afifi AA, Lachenbruch PA, Schouten HJA. Sample size determinations for the two-rater kappa statistic. *Psychometrika*. 1988;53:321–325.

Part II. Statistical Analysis Plan

1. Statistical Analysis Plan purpose and version

This Statistical Analysis Plan (SAP) prespecifies the statistical methods for the study entitled “Inter-Rater Agreement of Four Behavioral Rating Scales in Children Undergoing Dental Treatment Based on Clinical Video Recordings.” It is version 1.0, dated 04 September 2026, and is based on the protocol approved to the Human Research Ethics Committee of the Universitat de València (2026-ODON-4511724)

Any changes to this SAP after the start of data analysis will be documented, dated, justified, and identified as amendments.

2. Analysis objectives

The primary statistical objective is to quantify agreement among three evaluators applying four behavioral rating scales to the same video-recorded clinical events.

Secondary statistical objectives are to explore agreement by age, sex, treated tooth, and previous dental treatment experience when sample size permits.

3. Analysis populations

3.1 Full enrolled set

The full enrolled set will include all children who provide consent and are enrolled in the study.

3.2 Video-evaluable set

The primary agreement analysis will use the video-evaluable set, defined as enrolled participants with a recording of sufficient quality for at least one prespecified treatment stage and at least one rating from all three evaluators for the scale and stage being analyzed.

Agreement for a specific scale and stage will be calculated using all complete evaluator ratings available for that specific analysis. The number of included observations will be reported for every estimate.

3.3 Calibration set

The calibration set will consist of the five non-study cases used to assess evaluator calibration. These cases will not be included in the primary study analysis.

4. General statistical principles

All tests will be two-sided where hypothesis testing is appropriate. The nominal significance level will be 0.05. Because the principal purpose is estimation of inter-rater agreement rather than hypothesis testing, estimates will be presented with 95% confidence intervals and interpreted according to their magnitude and precision.

Continuous variables will be summarized using mean and standard deviation when approximately normally distributed and median, interquartile range, minimum, and maximum otherwise. Categorical variables will be summarized using counts and percentages.

5. Data preparation and quality checks

Before the statistical analysis, the database will be reviewed to identify duplicate codes, missing data, incorrect scores, and inconsistencies between the treatment stages and the available video recordings. The reasons why a video cannot be assessed will also be recorded. Any correction will be made only after checking the original record or a documented clarification. The final database will be closed before the main analysis begins.

6. Variable definitions

6.1 Participant variables

- Age: age in years at the time of recording.
- Sex: as recorded in the approved study dataset.
- Treated tooth: tooth treated during the recorded visit. Tooth identification using the FDI two-digit notation system

- Previous dental treatment experience: categorized according to the data-collection form.

6.2 Assessment variables

For every participant, treatment stage, scale, and evaluator, the corresponding behavioral category will be recorded.

The Houpt Scale will be analyzed by domain:

- Consciousness: 3 categories.
- Movement: 4 categories.
- Crying: 4 categories.
- Overall behavior: 6 categories.

OSUBRS Scale will be analyzed as a 4-category ordinal scale. Venham Scale will be analyzed as a 6-category ordinal scale and Frankl Scale will be analyzed as a 4-category ordinal scale.

7. Primary endpoint and analysis

7.1 Primary endpoint

The primary endpoint is inter-rater agreement among the three evaluators for behavioral scores obtained with each of the four scales during the prespecified treatment stages.

Because each scale has different domains and categories, the primary agreement result will be presented separately for each scale and stage rather than as a single pooled score.

7.2 Primary agreement statistic

For agreement among the three evaluators, Fleiss' kappa will be calculated for categorical ratings. For ordinal categories, a weighted multi-rater kappa or another prespecified ordinal agreement statistic will be used if available and appropriate in the selected statistical software.

If weighted multi-rater kappa cannot be reliably implemented in the software, unweighted Fleiss' kappa will be reported as the primary multi-rater statistic, and pairwise weighted kappa analyses will be reported as supplementary analyses. The selected implementation and weighting scheme will be documented in the statistical output.

Agreement estimates will be reported with 95% confidence intervals. The number of participants and ratings included in each analysis will be reported.

8. Secondary analyses

8.1 Pairwise evaluator agreement

Pairwise agreement will be estimated for Evaluator 1 versus Evaluator 2, Evaluator 1 versus Evaluator 3, and Evaluator 2 versus Evaluator 3 using Cohen's kappa or weighted Cohen's kappa, as appropriate.

8.2 Agreement by treatment stage

Agreement will be estimated separately for:

1. Reclining the dental chair.
2. Administration of infiltrative local anesthesia.
3. Placement of the rubber dam.
4. Use of rotary instruments.

8.3 Distribution of ratings

For each scale, stage, and evaluator, category counts and percentages will be presented. Agreement tables or cross-tabulations will be provided when useful for interpretation.

8.4 Exploratory subgroup analyses

If the data distribution permits, agreement will be explored by:

- Age group, (4–6 years versus 7–10 years)
- Sex.
- Previous dental treatment experience.
- Treated tooth or clinically relevant tooth category.

Subgroup analyses will be interpreted cautiously because the sample size is small and the study was not powered specifically for subgroup comparisons.

9. Missing data and non-evaluable recordings

No imputation will be performed for missing evaluator ratings or non-evaluable videos in the primary agreement analysis.

For each analysis, only observations with the required evaluator ratings will be included. The number and percentage of missing or non-evaluable observations will be summarized, together with the documented reason, such as technical recording failure, clinical interruption, insufficient image or audio quality, withdrawal of consent, or failure to capture the relevant treatment stage.

A sensitivity analysis may be performed if the pattern and amount of missing data could materially affect interpretation. Any sensitivity analysis will be identified as exploratory.

10. Sample size and precision

The target sample is 30 enrolled children. The study was designed to provide a practical and methodologically justified number of observations for estimating agreement across four scales and multiple treatment stages and for allowing possible losses due to unusable recordings or withdrawal.

The sample-size rationale uses $\alpha = 0.05$, 80% power, null $\kappa = 0.00$, alternative $\kappa = 0.60$, and an anticipated 15% allowance for losses. The formal minimum after this allowance was estimated as 18 participants, but the planned target was increased to 30 to improve stability, and increase representation of different behavioral categories.

Precision will be evaluated primarily through the width of 95% confidence intervals around κ estimates. No interim sample-size re-estimation is planned.

11. Software

Data may be entered and organized using Microsoft Excel version 16.101.1. Statistical analyses will be performed using IBM SPSS Statistics version 31.0.2.0 or another validated statistical package capable of producing the prespecified agreement statistics. The software version and procedures used for each analysis will be recorded.

12. Tables and reporting outputs

The statistical report will include, at minimum:

- Table 1: Participant demographic and clinical characteristics.
- Table 2: Number of evaluable recordings by treatment stage.
- Table 3: Distribution of ratings for each scale, stage, and evaluator.
- Table 4: Primary inter-rater agreement estimates with 95% confidence intervals.
- Table 5: Pairwise agreement estimates, if calculated.

- Supplementary tables: Subgroup and sensitivity analyses, if performed.

Results will be reported using aggregate data only. No table, figure, quotation, audio, or image will identify an individual child.