

Peer Review

Review of: "Agreement of Pain Assessment Using the Short Form of the Canine Glasgow Composite Measure Pain Scale between Veterinary Students, Veterinary Nurses, Veterinary Surgeons, and ECVAA-Diplomates"

Brunna Gomes Re¹

1. Independent researcher

Peer Review: Agreement of Pain Assessment Using the CMPS-SF Across Veterinary Roles

Article: Marco-Martorell M. et al., *Animals* (2024). DOI: <https://doi.org/10.3390/ani14162310>

Brief Summary

This prospective observational study evaluated agreement in CMPS-SF pain scoring across four rater groups: students, nurses or technicians, surgeons without formal anesthesia certification, and ECVAA diplomates, using 45 client-owned dogs assessed in vivo. Each assessment was video-recorded and later scored by 20 assessors (five per group). Overall inter-observer agreement was poor (ICC = 0.494). Within-group ICCs were highest for nurses (0.656), followed by ECVAA diplomates (0.540), surgeons (0.478), and students (0.432). The strongest inter-group agreement was observed between surgeons and nurses (ICC = 0.951) and between ECVAA diplomates and nurses (ICC = 0.951). The intervention threshold for additional analgesia was set at a CMPS-SF score of 6 or higher out of 24, and students were more likely to indicate rescue analgesia.

Strengths

- Clinically important question with direct implications for the timing of rescue analgesia
- Real-world cohort from a teaching hospital, which supports external validity

- Clear reliability framework using ICCs with comparisons across professional groups and an explicit intervention threshold

Major Comments

ICC specification and precision

Perhaps consider specifying the exact ICC models used (two-way random versus mixed, absolute agreement versus consistency, single-measure versus average-measure) and presenting 95 percent confidence intervals for all ICCs. Single-measure ICCs are most relevant for day-to-day clinical use. A Bland–Altman analysis for key group pairs would help visualize bias and limits of agreement.

Score distribution and case mix

Reliability is sensitive to variance in true pain levels. Providing a distribution of total CMPS-SF scores (histogram or violin plot) and describing the case mix (procedure types, timing post-operative, sedation, and analgesic protocols) would contextualize ICC values. A restricted score range can depress ICCs; documenting spread will aid interpretation.

Video protocol and item coverage

Because CMPS-SF items rely on standardized observation and handling, a concise summary of the video protocol would improve reproducibility. Please clarify who applied stimuli, clip duration, camera angles, sequence of observations, and whether any items were not assessable from video. Briefly discuss how these constraints might influence agreement.

Training and calibration

The higher frequency of rescue calls by students suggests a calibration gap. It would be helpful to clarify any pre-study training and to discuss the feasibility of a brief, structured calibration module of 10 to 15 minutes using annotated exemplars for all raters. A small pre–post pilot reporting change in ICC and change in rescue-triggering rates would provide actionable guidance.

Decision accuracy near the rescue threshold

ICCs quantify agreement but not classification accuracy. Receiver operating characteristic analyses using a diplomate consensus as the reference could estimate sensitivity and specificity for rescue needed, particularly near the cut-off. Operationally, a policy such as obtaining a second rater when scores are near the threshold could mitigate misclassification.

Generalizability and minimum competency

It would complement the results if the reported student training stage, nurse or technician years in

practice, and prior exposure to pain scales for surgeons were included. Offering a minimal competency package, for example, six annotated cases plus a quick-reference card, could facilitate translation beyond teaching hospitals.

Minor Comments

It would bring wider results to clarify blinding to clinical context and to administered analgesia at the time of scoring, and to specify the time points when assessments were performed.

Maintaining consistent CMPS-SF terminology and citing the primary user guide for rescue thresholds would be recommended.

Add a table with mean plus or minus standard deviation CMPS-SF totals by group and list all ICCs with 95 percent confidence intervals; compact violin or box plots would help readers visualize score distributions and overlap.

Clinical Implications

The combination of modest single-rater reliability and stronger agreement of group averages supports three low-cost steps in practice:

- Brief initial calibration for all staff.
- Automatic second opinion when scores fall near the rescue threshold.
- Periodic re-calibration using annotated exemplars to sustain consistency and animal welfare.

Ethics and Data

Consider sharing de-identified exemplar clips and aggregated scoring sheets in a suitable repository to enable training and replication.

Conflicts of Interest

None declared.

ORCID: <https://orcid.org/0009-0000-9832-0704>

Declarations

Potential competing interests: No potential competing interests to declare.