



Construct validity, responsiveness and reliability of the Feline Grimace Scale in kittens

Alice J Cheng¹ , Annie Malo¹ , Marta Garbin¹ ,
Beatriz P Monteiro¹ and Paulo V Steagall^{1,2}

Journal of Feline Medicine and Surgery
1–8

© The Author(s) 2023

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1098612X231211765

journals.sagepub.com/home/jfm

This paper was handled and processed by the European Editorial Office (ISFM) for publication in *JFMS*



Abstract

Objectives The aim of the present study was to investigate the construct validity, responsiveness and reliability of the Feline Grimace Scale (FGS) in kittens.

Methods A total of 36 healthy female kittens (aged 10 weeks to 6 months) were included in a prospective, randomized, blinded study. Video recordings of all kittens were made before and 1 and 2 h after ovariohysterectomy using an opioid-free injectable anesthetic protocol with or without multimodal analgesia. Additional recordings were taken before and 1 h after administration of rescue analgesia (buprenorphine 0.02 mg/kg IM) to painful kittens. Screenshots of facial images were collected from the video recordings for FGS scoring. Four observers blinded to treatment groups and time points scored 111 randomized images twice with a 5-week interval using the FGS. Five action units (AUs) were scored (ear position, orbital tightening, muzzle tension, whiskers position and head position; 0–2 each). Construct validity, responsiveness, and inter- and intra-rater reliability were evaluated using linear models with Benjamini–Hochberg correction, Wilcoxon signed-rank test and single intraclass correlation coefficients (ICC_{single}), respectively ($P < 0.05$).

Results FGS total ratio scores were higher at 1 and 2 h after ovariohysterectomy (median [interquartile range, IQR]: 0.30 [0.20–0.40] and 0.30 [0.20–0.40], respectively) than at baseline (median [IQR]: 0.10 [0.00–0.30]) ($P < 0.001$). FGS total ratio scores were lower after the administration of rescue analgesia (median [IQR] before and after rescue analgesia) 0.40 [0.20–0.50] and 0.20 [0.10–0.38], respectively ($P < 0.001$). Inter-rater ICC_{single} was 0.68 for the FGS total ratio scores and 0.35–0.70 for all AUs considered individually. Intra-rater ICC_{single} was 0.77–0.91 for the FGS total ratio scores and 0.55–1.00 for all AUs considered individually.

Conclusions and relevance The FGS is a valid and responsive acute pain-scoring instrument with moderate inter-rater reliability and good to excellent intra-rater reliability in kittens.

Keywords: Analgesia; kitten; pain; pain assessment; facial expressions; Feline Grimace Scale

Accepted: 13 October 2023

Introduction

There have been significant advances in feline peri-operative pain management with the publication of three instruments for acute pain assessment:¹ the Glasgow Composite Measure Pain Scale-Feline (CMPS-F);² the UNESP-Botucatu multidimensional feline pain assessment scale (UFEPS) and its short form (UFEPS-SF),^{3,4} and the Feline Grimace Scale (FGS).^{5–7} The FGS is a facial expression-based scoring system comprising five action units (AUs): ear position, orbital tightening, muzzle tension, whiskers position and head position. Each AU is scored from 0 to 2, for a maximum FGS score of 10. Total ratio scores are obtained by considering the sum of scores

¹Department of Clinical Sciences, Faculty of Veterinary Medicine, Université de Montréal, Saint-Hyacinthe, QC, Canada

²Department of Veterinary Clinical Sciences, Centre for Companion Animal Health and Welfare, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong SAR, China

Corresponding author:

Paulo V Steagall DVM, MSc, PhD, Diplomate ACVAA, Department of Veterinary Clinical Sciences, Centre for Companion Animal Health and Welfare, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong SAR, China

Email: pmortens@cityu.edu.hk



for each visible AU divided by the maximum possible score. An FGS total ratio score $\geq 0.39/1$ suggests that analgesia should be administered.⁵ The FGS has demonstrated validity and reliability in various acute painful conditions,^{5,8} when evaluated using real-time or image assessment,⁶ and when used by individuals with differing levels of expertise, including cat caregivers.^{9,10}

Although some of these assessment tools have gone through extensive validation in adult cats, none have been uniquely validated in kittens. Nonetheless, early-age neutering (≤ 4 months of age) is now common practice in feline medicine and is supported by organizations such as the International Society of Feline Medicine for optimal health, social development and population control.^{11,12} Considering that thousands of kittens undergo early-age neutering every year, it is concerning that pain assessment tools have yet not been validated in this population to ensure that pain is recognized and proper analgesic treatment is administered based on objective pain scores.

The objective of the present study was to investigate the construct validity, responsiveness, and inter- and intra-rater reliability of the FGS in kittens undergoing ovariohysterectomy (OVH). The hypothesis was that the FGS total ratio scores would be significantly different before and after surgery, and before and after the administration of analgesics, and would present good reliability between raters and within raters over time.

Materials and methods

Ethical statement

The present study was conducted in parallel with a prospective, randomized, blinded clinical trial comparing an opioid-free injectable anesthetic protocol with or without multimodal analgesia in kittens undergoing OVH.¹³ The study protocol was approved by the Comité d'éthique de l'utilisation des animaux (Animal Use Ethics Committee) of the Faculty of Veterinary Medicine, Université de Montréal (protocol 21-Rech-2132), performed in accordance with the Canadian Council on Animal Care and reported according to the ARRIVE guidelines.¹⁴

Animals

A total of 40 domestic healthy female kittens of any breed, aged between 10 weeks and 6 months, weighing ≥ 1 kg and with a body condition score of 4–6/9 were recruited by one of the investigators (AJC) from two local shelters between July and August 2021. Kittens were excluded if they were beyond the age limits, male, or showing shy or feral behaviours or signs of disease (pain, hyperthermia, etc.). Any other individual was included in the study. Kittens were admitted to the veterinary teaching hospital (Centre hospitalier universitaire vétérinaire) of the Faculty of Veterinary Medicine, Université de Montréal approximately 16 h before elective OVH. Written consent was obtained for each kitten.

After admission, kittens were individually and randomly housed in adjacent stainless-steel cages in a cat ward with temperature control. Each kitten had access to water ad libitum, a litter box, a blanket, a cardboard box and toys, except during filming periods (see 'Pain assessment and video recording'). Kittens were left undisturbed for 3 h in their cages to acclimatize to the study site and personnel. Thereafter, a detailed physical examination was performed by two veterinarians (AM, BPM) using low-stress, feline-friendly interactive techniques.^{15,16} Soft food (Royal Canin Gastrointestinal Kitten; Royal Canin) was offered at specific time points in proportion to each kitten's individual daily calorie requirements. Kittens were discharged 24 h after surgery and returned to their respective shelters for adoption.

Anesthesia and surgery

Briefly, food was removed 6–10 h before the induction of general anesthesia, but water remained accessible. Each kitten received 0.5 ml of corn syrup applied directly to the gingival membrane on the morning of surgery to prevent hypoglycemia. All kittens received an intramuscular (IM) injection of ketamine (4 mg/kg; Ketaset; Zoetis), dexmedetomidine (40 μ g/kg; Dexdomitor; Zoetis) and midazolam (0.25 mg/kg; Midazolam; Sandoz) administered together into the lumbar epaxial muscles using a 1 ml syringe. A randomization plan assigned each kitten to either the multimodal group (MMG) or control group (CG). In the MMG, animals received subcutaneous (SC) meloxicam (0.1 mg/kg; Metacam 0.5%; Boehringer Ingelheim) before surgery and bupivacaine hydrochloride 0.25% (2 mg/kg; Bupivacaine injection BP; Sterimax) administered intraperitoneally before OVH. In the CG, kittens received saline (0.9% sodium chloride solution) at equal volumes and routes of administration as the meloxicam and bupivacaine in the MMG. The same veterinarian with experience in surgery (BPM) performed an elective OVH using the pedicle tie technique. During anesthetic recovery, kittens received 5 ml/kg of lactated Ringer's solution (Lactated Ringer's Inj. Bag/500 ml; McCarthy & Sons Service) SC between the shoulder blades. An IM injection of atipamezole (0.4 mg/kg; Antisedan; Zoetis) was administered 15 mins after the end of surgery and each kitten was returned to its respective cage for pain assessment and postoperative care.

Pain assessment and video recording

Pain was assessed in real time using the UFEPS-SF before surgery (baseline) and 1, 2, 4, 6, 8, 12 and 24 h postoperatively. Pain assessment was performed by a female veterinarian (AM) who was blinded to the treatment groups. This individual had 7 years of experience in clinical practice. Before the study, she completed training using the UFEPS with videos available on the website www.animalpain.org. Rescue analgesia was administered to kittens with UFEPS-SF scores $\geq 4/12$ using

IM buprenorphine (0.02 mg/kg; Vetergesic; Champion Alstoe) in the MMG and CG, and SC meloxicam (0.1 mg/kg) in the CG. The administration of the rescue analgesia was performed by veterinarians who were not involved with pain assessment. Regardless of the treatment group, kittens weighing over 2 kg or older than 16 weeks received a second dose of meloxicam by the oral route of administration (0.05 mg/kg; Metacam 0.5 mg/ml oral suspension; Boehringer Ingelheim) 24 h after the first dose.

Video recordings of the kittens were made by an observer (AJC) using high-definition wide-angle cameras (GoPro Hero 5 and GoPro Hero 9; GoPro) at 60 frames per second. Cameras were attached to the cage bars at the kitten's eye level and adjusted when needed. The cage walls were covered with a light-colored cardboard to ensure background uniformity during filming and image assessment. Diffused external lights were used for appropriate lighting. Cages were emptied apart from a blanket and the kitten. During FGS video recording, kittens were left undisturbed. Videos with a duration of 3 mins were recorded at baseline, 1 and 2 h after OVH, and before and 1 h after the administration of rescue analgesia. In some cases, extended videos (total of 9 mins) were collected for the purpose of a future study involving kitten behavior, but only the first 3 mins were used for the purpose of this study.

Image capture, selection and cropping

A conversion software (Free Video to JPG Converter; DVDVideoSoft) was used to collect still images from video recordings with the output set at two images per second. An observer (AJC) who was blinded to the treatment group and UFEPS-SF pain scores selected one image for each third of the video, collecting three images per video in total. For an image to be included in the study, the kitten had to be facing the camera with visible AUs. If it was leaning on to a surface (eg, the cage wall or floor), the image could still be included if the AUs were visible on the contralateral side. Images were excluded if they were of poor quality or if kittens were vocalizing, grooming or sleeping. Of the three images selected per video, the best image (in terms of quality and definition) was chosen to characterize the kitten at that time point for later scoring. Subsequently, images were cropped to include the entire face of the kitten using software (Adobe Photoshop CS6 V13.0; Adobe). Brightness was increased for dark images. A short video is provided, illustrating the procedure for video recording, image capturing, selection and cropping (see video in supplementary material).

Image scoring

Four raters (AM, BPM, MG and PVS) received the FGS training manual and participated in a training session provided by the principal investigator (PVS) before image scoring.⁵ In this session, raters independently scored 10

images from the 'Practice your Skills' menu of the FGS website (www.felinegrimacescale.com) and discussed their scores. Thereafter, an online survey (LimeSurvey; LimeSurvey GmbH) was built to allow scoring of the final selected images. Images were randomly ordered using a random sequence generator (www.random.org). For each image, the rater, blinded to treatment groups and time points, scored the five AUs of the FGS from 0 to 2: 0 = AU is absent; 1 = moderate appearance of the AU or uncertainty over its presence or absence; and 2 = obvious appearance of the AU. If the AU was not clearly visible, raters had the option of checking 'not possible to score'. An image was excluded from statistical analysis when two or more AUs were not possible to score by at least one rater. An FGS total ratio score for each image scored by each rater was calculated by summing the scores from each AU and dividing by the maximum possible score, excluding the AUs marked as 'not possible to score'. Image scoring was divided into two parts; raters were instructed to complete evaluations at least 24 h apart to avoid fatigue. Five weeks later, the same procedure was repeated with raters scoring the same images using a different randomized sequence. Image scoring was performed between November 2021 and January 2022.

Statistical analysis

Statistical analyses were performed using R (R Core Team V4.0.3; R Foundation for Statistical Computing). Construct validity was assessed by comparing FGS total ratio scores at baseline vs 1 and 2 h. A repeated measures linear mixed model with the Benjamini–Hochberg correction was used. The kitten was treated as a random effect and time as fixed effect. Responsiveness was assessed by comparing FGS total ratio scores before and 1 h after the administration of rescue analgesia. Normality was verified with the Kolmogorov–Smirnov test and through visual inspection. The FGS total ratio scores were not normally distributed and the Wilcoxon signed-rank test was used for group comparisons. Inter- and intra-rater reliability were assessed for each AU and for the FGS total ratio scores using the intraclass correlation coefficient (ICC) with the 95% confidence interval (CI). Inter-rater reliability was assessed using a two-way random effect model for absolute agreement, whereas intra-rater reliability was assessed using a two-way mixed effect model. The single and average ICCs (ICC_{single} and ICC_{average}, respectively) were interpreted as follows: <0.5 = poor reliability; 0.5–0.75 = moderate reliability; 0.75–0.9 = good reliability; and >0.90 = excellent reliability.¹⁷ *P* values <0.05 were considered statistically significant.

Results

A total of 111 images of 36 female kittens (mean age 15 ± 4.7 weeks; mean body weight 1.5 ± 0.4 kg) were included in the statistical analysis (Figure 1). The mean \pm

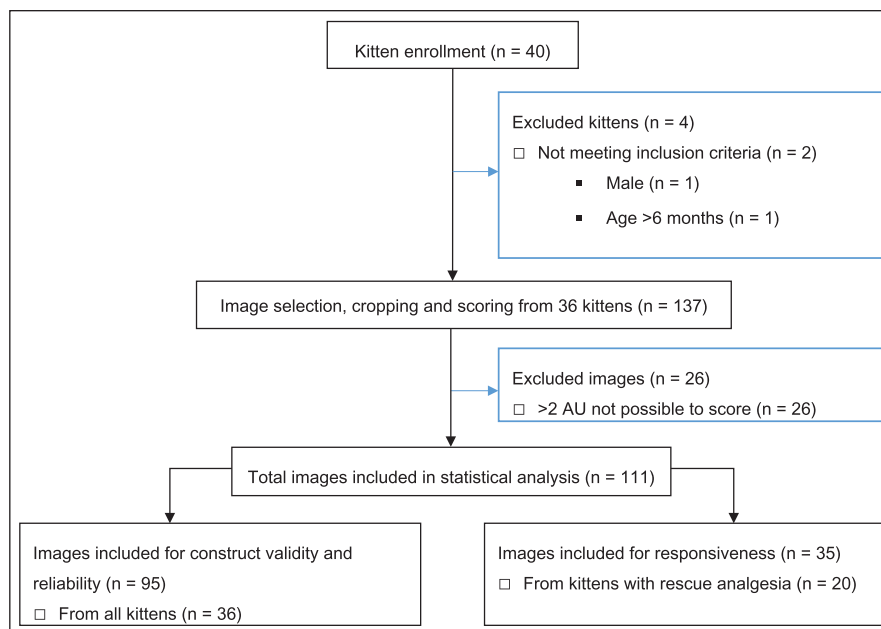


Figure 1 Flow chart of a prospective, randomized and blinded study investigating the construct validity, responsiveness, and inter-rater and intra-rater reliability of the Feline Grimace Scale using facial images of kittens undergoing ovariohysterectomy

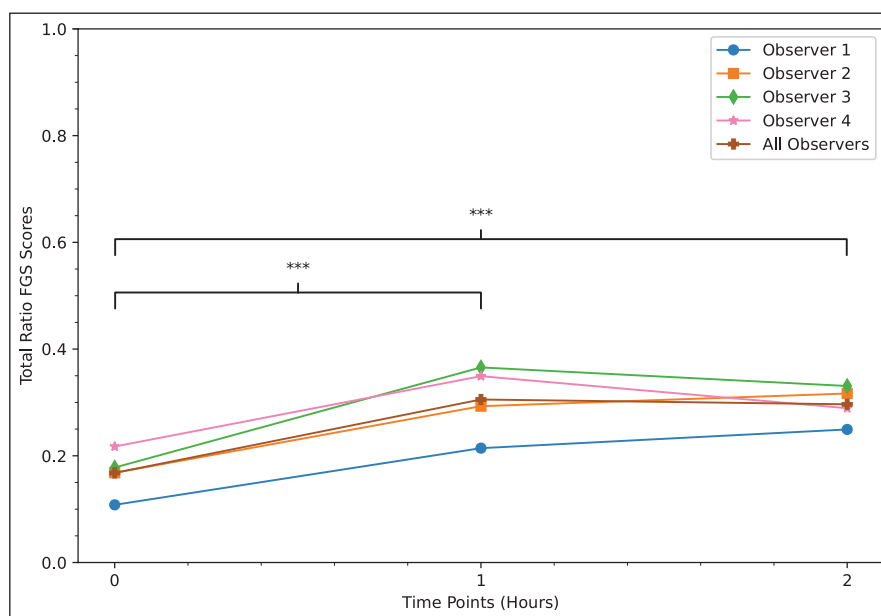


Figure 2 Median total ratio Feline Grimace Scale (FGS) scores in kittens assessed using known-group discrimination. There were significant increases of the FGS total ratio scores at 1 and 2h after surgery compared with baseline ($P < 0.001$), which demonstrates construct validity

standard deviation for the duration of FGS image scoring was 73 ± 33 mins and 68 ± 23 mins for the first and second round of assessments, respectively.

A total of 95 images of 36 kittens were used to assess construct validity by known-group discrimination. The FGS total ratio scores were increased at 1 and

2h (median [interquartile range, IQR range]: 0.30 [0.20–0.40] and 0.30 [0.20–0.40], respectively) when compared with baseline (median [IQR]: 0.10 [0.00–0.30]) ($P < 0.001$) (Figures 2 and 3).

A total of 35 images of 20 kittens were used to assess responsiveness. The FGS total ratio scores were

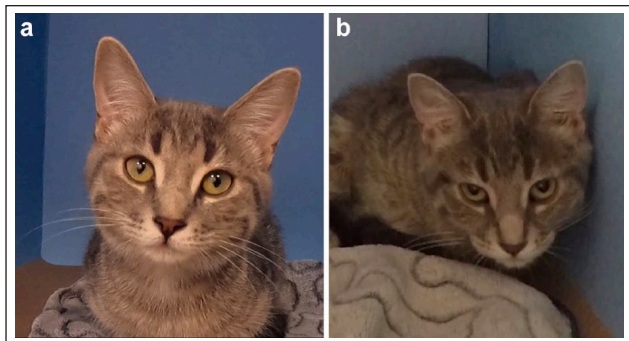


Figure 3 An example of a kitten with changes in Feline Grimace Scale total ratio scores (a) before and (b) after surgery, demonstrating the construct validity of the tool

significantly decreased 1 h after administration of rescue analgesia (median [IQR] before and after rescue: 0.40 [0.20–0.50] and 0.20 [0.10–0.38], respectively) ($P < 0.001$) (Figures 4 and 5).

A total of 95 images of 36 kittens were used to assess reliability. The inter-rater reliability of the FGS total ratio scores was moderate and excellent according to the ICC_{single} and ICC_{average} , respectively. The ICC_{single} was moderate for ear position, orbital tightening and head position, and poor for muzzle tension and whiskers position. The ICC_{average} was excellent for orbital tightening, good for ear position, muzzle tension and head position, and moderate for whiskers position (Table 1). The intra-rater reliability for the FGS total ratio scores of the four raters was good to excellent. The ICC_{single} ranged from

moderate to excellent for ear position and orbital tightening, moderate to good for muzzle tension, moderate for whiskers position, and good to excellent for head position (Table 2).

Discussion

In the present study, the validation of the FGS in kittens is reported by evaluating its construct validity, responsiveness and reliability using image assessment. Overall, the results indicated that the inter-rater reliability was moderate using ICC_{single} and excellent using ICC_{average} , whereas intra-rater reliability ranged from good to excellent for the FGS total ratio scores. Similar to what is observed in adult cats, the FGS total ratio scores increased after surgery and decreased after the administration of rescue analgesia.⁵ These results are important for feline health and welfare, as the FGS is now the first acute pain assessment tool with reported validity specifically in kittens. With the widespread adoption of early-age spaying in cats, the tool can now be used in high-volume, high-quality spay/neuter programs to ensure pain recognition and treatment based on a more objective assessment.

Construct validity was assessed using known-groups discrimination by comparing the FGS total ratio scores before surgery (non-painful) vs after surgery (painful). The hypothesis was that kittens before surgery would have lower FGS total ratio scores than after surgery. Comparisons were limited to baseline vs 1 and 2 h, when acute pain is suspected to be most significant after OVH and when rescue analgesia is commonly required based on our previous studies.^{5,6,8,18} The FGS discriminated

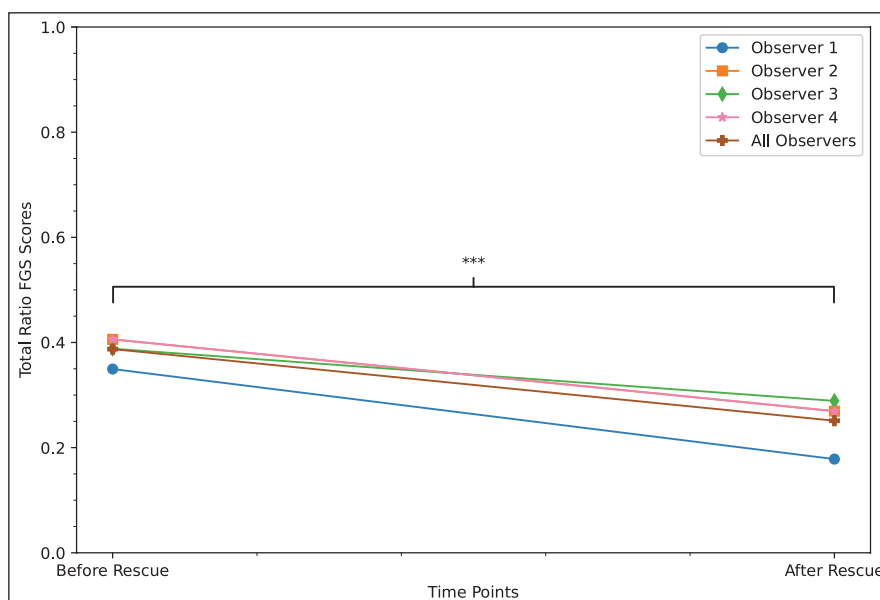


Figure 4 Median total ratio Feline Grimace Scale (FGS) scores after an intramuscular injection of buprenorphine with or without subcutaneous meloxicam in kittens showing signs of pain after ovariohysterectomy. There were significant decreases of the FGS total ratio scores 1 h after rescue analgesia ($P < 0.001$), demonstrating responsiveness of the FGS in kittens

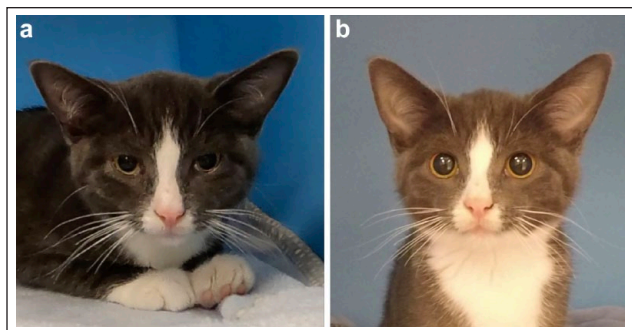


Figure 5 An example of a kitten with changes in Feline Grimace Scale total ratio scores (a) before and (b) after the administration of buprenorphine and meloxicam by the intramuscular and subcutaneous routes, respectively, demonstrating the responsiveness of the tool

non-painful and painful kittens, as scores were significantly increased at 1 and 2 h when compared with baseline. Similar results were observed in adult cats in a previous study: client-owned cats with painful conditions had higher FGS scores than non-painful control cats.⁵

Responsiveness consists of the ability of an instrument to detect clinically important changes over time in the construct measured.^{7,19,20} In this study, the discrimination between the FGS total ratio scores before and 1 h after analgesic intervention was assessed. The intervention in this study consisted of administration of buprenorphine with or without meloxicam for rescue analgesia when kittens received UFEPS-SF pain scores $\geq 4/12$. Significant decreases in FGS total ratio scores were observed 1 h after the administration of rescue analgesia to kittens in pain. Thus, the FGS is responsive to analgesic treatment in kittens undergoing OVH. Similar results were observed in adult cats with medical or surgical pain after the administration of analgesia.^{5,6}

Reliability describes the capacity of measurements to be inherently reproducible.²¹ The measurement can be tested between raters (inter-rater reliability) and over time for the same rater (intra-rater reliability) using the same outcomes (ie, images, videos, etc). The inter-rater

Table 1 Inter-rater reliability of the FGS in kittens undergoing ovariohysterectomy

Action unit	ICC _{single} (95% CI)	ICC _{average} (95% CI)
Ear position	0.62 (0.53–0.71)	0.77 (0.61–0.85)
Orbital tightening	0.70 (0.61–0.77)	0.92 (0.89–0.95)
Muzzle tension	0.47 (0.36–0.59)	0.83 (0.68–0.90)
Whiskers position	0.35 (0.22–0.48)	0.63 (0.42–0.77)
Head position	0.66 (0.53–0.76)	0.89 (0.83–0.93)
FGS total ratio scores	0.68 (0.57–0.77)	0.91 (0.87–0.94)

Data were obtained from round 1 assessments and are reported as single and average intraclass correlation coefficient (ICC_{single} and ICC_{average}, respectively) and 95% confidence interval (CI). The ICC_{single} and ICC_{average} were interpreted as follows: <0.5 = poor reliability; 0.5 – 0.75 = moderate reliability; 0.75 – 0.9 = good reliability; and >0.90 = excellent reliability²¹
FGS = Feline Grimace Scale

reliability (ICC_{single}) of the FGS total ratio scores was moderate in this study. Previous studies showed good inter-rater reliability using the FGS in adult cats.^{5,8,10} The reliability of whiskers position was relatively lower compared with the other AUs, which is also similar to previous findings.^{5,8} The interpretation of this AU could be difficult when assessment is performed using an image, as it can be affected by camera position, lighting and image quality. It is also important to note that the 95% CI (ie, 95% chance that the true ICC value will be at any point within the CI for that AU) includes values for which the ICC interpretation could change. In addition, the ICC_{average}, an index for the reliability of mean of k raters, revealed moderate inter-rater reliability for whiskers position, good inter-rater reliability for ear position, muzzle tension and head position, and excellent inter-rater reliability for orbital tightening and FGS total ratio scores, according to the classification used. Comparisons among studies are not easy as different interpretations and calculations can be used for ICC, and it is not always known what ICC type (ie, single or average) is actually reported in studies.⁷ In our study, ICC calculations were performed using

Table 2 Intra-rater reliability of the FGS scores in kittens undergoing ovariohysterectomy

Action unit	Rater 1	Rater 2	Rater 3	Rater 4
Ear position	0.71 (0.59–0.80)	0.75 (0.64–0.82)	0.65 (0.52–0.76)	0.93 (0.90–0.95)
Orbital tightening	0.69 (0.57–0.78)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.92 (0.88–0.95)
Muzzle tension	0.70 (0.58–0.79)	0.75 (0.64–0.82)	0.68 (0.55–0.77)	0.86 (0.79–0.90)
Whiskers position	0.55 (0.39–0.68)	0.57 (0.42–0.69)	0.66 (0.53–0.76)	0.75 (0.65–0.83)
Head position	0.89 (0.85–0.93)	0.79 (0.69–0.85)	0.78 (0.68–0.85)	0.95 (0.92–0.97)
FGS total ratio scores	0.82 (0.75–0.88)	0.78 (0.69–0.85)	0.77 (0.67–0.84)	0.91 (0.86–0.94)

Raters scored each image twice in a different randomized order 5 weeks apart. Data are reported as single intraclass correlation coefficient (ICC_{single}) and 95% confidence interval (CI). The ICC_{single} was interpreted as follows: <0.5 = poor reliability; 0.5 – 0.75 = moderate reliability; 0.75 – 0.9 = good reliability; and >0.90 = excellent reliability²¹
FGS = Feline Grimace Scale

two-way random effect models because the results can be generalized to any raters who possess the same characteristics as the selected raters in our study (ie, veterinarians). Absolute agreement, and not consistency, was selected for the ICC definition, as it is not expected that the raters' scores would be correlated in an additive manner.^{17,22} The ICC_{average} is usually higher than the ICC_{single} and it can often be used to 'inflate the results'. However, the ICC_{single} is best selected when the basis of the actual clinical measurement (ie, FGS scoring) involves a single rater during its application.¹⁷ Finally, our classification for ICC values is more conservative than the one used by Altman (<0.2=poor; 0.21–0.4=reasonable; 0.41–0.60=moderate; 0.61–0.80=good; and 0.81–1.0=very good),²³ which may contribute to different interpretations of our results and the use of the FGS in the clinical setting.⁹

The intra-rater reliability of the FGS total ratio scores in kittens was good to excellent, which was similar to previous studies with images from adult cats, when ICCs from the FGS total ratio scores from veterinarians were excellent.^{5,10} Orbital tightening and ear position ranged from moderate to excellent, muzzle tension ranged from moderate to good, whiskers position was moderate and head position ranged from good to excellent. The intra-rater reliability of veterinarians for individual AUs was evaluated in adult cats using the same method for classification.^{10,17} The ICC classification between this previous and the present study was similar for all AUs, except for head position, which ranged from good to excellent in the present study and was moderate in the previous study. While knowing the inter- and intra-rater reliability of each AU is important in research, the FGS score is considered exclusively in the clinical setting as it informs the need for analgesic treatment based on the interventional score. Similar results were reported in adult cats for the FGS total ratio scores and for each AU, but with altogether higher reliability than in kittens.^{5,8,10} This was surprising as raters participated in a training session before the study began. As much as the effects of training on FGS scoring are not known, we expected that our results would be similar to previous studies using the FGS in adult cats. This difference could be because raters may be less familiar with kittens than with adult cats.

The present study has some limitations. Image quality could still be improved. Despite efforts to add external lighting, a similar background and to use a better camera than previous FGS studies,^{5,8} the process of image selection and cropping still led to the exclusion of many images due to poor quality. Suboptimal images may have had greater impact when scoring fine AUs, such as whiskers position and muzzle tension, than when scoring head position, ear position and orbital tightening; the latter may be easier to discern than the former ones. It is important to consider that the construct validity, responsiveness and reliability of the FGS were studied

only in female kittens undergoing OVH. As much as this is certainly the target population considering early-age spaying practices, it is not known if similar results would have been obtained with medical pain, other types of surgeries or trauma, and with kittens presenting feral or shy behaviors. The exclusion of behavior-specific populations represents a potential bias to the validation and application of the FGS in kittens, as spay/neuter programs may include individuals with all types of behavioral traits. Responsiveness of the FGS was not tested for control, non-painful kittens, to demonstrate that scores would not change over time for these individuals in a hospital setting. On the other hand, data were robust, including a heterogeneous set of images of both painful and pain-free kittens for comparisons.

Conclusions

The FGS demonstrated high discriminative ability and responsiveness to an analgesic treatment with overall moderate inter-rater reliability, and good to excellent intra-rater reliability, in kittens. The FGS is a valid and reliable instrument for acute pain assessment in kittens, representing a substantial improvement for feline pain management in the context of spay/neuter programs.

Acknowledgements We are thankful to all co-authors of the clinical trial who were involved in video recordings and Sean MacRae for manuscript revision and plotting.

Author note An abstract of this study was part of an oral presentation at the 2022 Nafplio Association of Veterinary Anaesthetists Spring Meeting. This manuscript represents a portion of an MSc degree thesis of the first author to the Université de Montréal.

Supplementary material The following files are available as supplementary material:
Supplementary material video
ARRIVE Checklist.

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

Funding Paulo V Steagall's laboratory is funded by a Discovery Grant of the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-03831).


Ethical approval The work described in this manuscript involved the use of non-experimental (owned or unowned) animals. Established internationally recognized high standards ('best practice') of veterinary clinical care for the individual patient were always followed and/or this work involved the use of cadavers. Ethical approval from a committee was therefore not specifically required for publication in *JFMS*. Although not required, where ethical approval was still obtained, it is stated in the manuscript.


Informed consent Informed consent (verbal or written) was obtained from the owner or legal custodian of all animal(s) described in this work (experimental or non-experimental animals, including cadavers) for all procedure(s) undertaken (prospective or retrospective studies). For any animals or people individually identifiable within this publication, informed consent (verbal or written) for their use in the publication was obtained from the people involved.

ORCID iD Alice J Cheng  <https://orcid.org/0009-0007-2426-5459>

Annie Malo  <https://orcid.org/0000-0001-7692-9527>

Marta Garbin  <https://orcid.org/0000-0002-8156-3950>

Beatriz P Monteiro  <https://orcid.org/0000-0002-5722-5687>

Paulo V Steagall  <https://orcid.org/0000-0003-4150-6043>

References

- 1 Steagall PV and Monteiro BP. **Acute pain in cats: recent advances in clinical assessment.** *J Feline Med Surg* 2019; 21: 25–34.
- 2 Reid J, Scott EM, Calvo G, et al. **Definitive Glasgow acute pain scale for cats: validation and intervention level.** *Vet Rec* 2017; 180: 449. DOI: 10.1136/vr.104208.
- 3 Belli M, de Oliveira AR, de Lima MT, et al. **Clinical validation of the short and long UNESP-Botucatu scales for feline pain assessment.** *PeerJ* 2021; 9. DOI: 10.7717/peerj.11225.
- 4 Luna SPL, Trindade PHE, Monteiro BP, et al. **Multilingual validation of the short form of the Unesp-Botucatu Feline Pain Scale (UFEPS-SF).** *PeerJ* 2022; 10. DOI: 10.7717/peerj.13134.
- 5 Evangelista MC, Watanabe R, Leung VSY, et al. **Facial expressions of pain in cats: the development and validation of a Feline Grimace Scale.** *Sci Rep* 2019; 9: 19128. DOI: 10.1038/s41598-019-55693-8.
- 6 Evangelista MC, Benito J, Monteiro BP, et al. **Clinical applicability of the Feline Grimace Scale: real-time versus image scoring and the influence of sedation and surgery.** *PeerJ* 2020; 8. DOI: 10.7717/peerj.8967.
- 7 Evangelista MC, Monteiro BP and Steagall PV. **Measurement properties of grimace scales for pain assessment in nonhuman mammals: a systematic review.** *Pain* 2022; 163: e697–e714.
- 8 Watanabe R, Doodnaught GM, Evangelista MC, et al. **Inter-rater reliability of the feline grimace scale in cats undergoing dental extractions.** *Front Vet Sci* 2020; 7. DOI: 10.3389/fvets.2020.00302.
- 9 Monteiro BP, Lee NH and Steagall PV. **Can cat caregivers reliably assess acute pain in cats using the Feline Grimace Scale? A large bilingual global survey.** *J Feline Med Surg* 2023; 25. DOI: 10.1177/1098612x221145499.
- 10 Evangelista MC and Steagall PV. **Agreement and reliability of the Feline Grimace Scale among cat owners, veterinarians, veterinary students and nurses.** *Sci Rep* 2021; 11: 5262. DOI: 10.1038/s41598-021-84696-7.
- 11 Sparkes AH, Bessant C, Cope K, et al. **ISFM guidelines on population management and welfare of unowned domestic cats (*Felis catus*).** *J Feline Med Surg* 2013; 15: 811–817.
- 12 Joyce A and Yates D. **Help stop teenage pregnancy! Early-age neutering in cats.** *J Feline Med Surg* 2011; 13: 3–10.
- 13 Malo A, Cheng AJ, Ruel HL, et al. **Randomised, prospective, blinded, clinical trial of opioid-free injectable anaesthesia with or without multimodal analgesia in kittens undergoing ovariohysterectomy.** *J Feline Med Surg* 2023; 25. DOI: 10.1177/1098612x231158582.
- 14 Percie du Sert N, Hurst V, Ahluwalia A, et al. **The ARRIVE guidelines 2.0: updated guidelines for reporting animal research.** *PLoS Biol* 2020; 18. DOI: 10.1371/journal.pbio.3000410.
- 15 Rodan I, Dowgray N, Carney HC, et al. **2022 AAFP/ISFM cat friendly veterinary interaction guidelines: approach and handling techniques.** *J Feline Med Surg* 2022; 24: 1093–1132.
- 16 Rodan I, Sundahl E, Carney H, et al. **AAFP and ISFM feline-friendly handling guidelines.** *J Feline Med Surg* 2011; 13: 364–375.
- 17 Koo TK and Li MY. **A guideline of selecting and reporting intraclass correlation coefficients for reliability research.** *J Chiropr Med* 2016; 15: 155–163.
- 18 Grint NJ, Murison PJ, Coe RJ, et al. **Assessment of the influence of surgical technique on postoperative pain and wound tenderness in cats following ovariohysterectomy.** *J Feline Med Surg* 2006; 8: 15–21.
- 19 Husted JA, Cook RJ, Farewell VT, et al. **Methods for assessing responsiveness: a critical review and recommendations.** *J Clin Epidemiol* 2000; 53: 459–468.
- 20 Guyatt GH, Deyo RA, Charlson M, et al. **Responsiveness and validity in health status measurement: a clarification.** *J Clin Epidemiol* 1989; 42: 403–408.
- 21 Lachin JM. **The role of measurement reliability in clinical trials.** *Clin Trials* 2004; 1: 553–566.
- 22 McGraw KO and Wong SP. **Forming inferences about some intraclass correlation coefficients.** *Psychological Methods* 1996; 1: 30–46.
- 23 Altman DG. **Practical statistics for medical research.** London: Chapman and Hall, 1991.