314 (2025) 106448

Contents lists available at ScienceDirect

The Veterinary Journal

journal homepage: www.elsevier.com/locate/tvjl

Understanding the Feline Grimace Scale: A study of dimensional structure, importance of each action unit and variables affecting assessment

Paulo V. Steagall [a,b,c,*], Beatriz P. Monteiro [a], Pedro H.E. Trindade [d,e], Syed S.U.H. Bukhari [b,c], Stelio P.L. Luna [e]

[a] Department of Clinical Sciences, Faculty of Veterinary Medicine, Université de Montréal, 3200 rue Sicotte, Saint-Hyacinthe, QC J2S 2M2, Canada
[b] Department of Veterinary Clinical Sciences, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong Special Administrative Region of China
[c] Centre for Animal Health and Welfare, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong Special Administrative Region of China
[d] Department of Large Animal Clinical Sciences, College of Veterinary Medicine, Michigan State University (MSU), East Lansing, MI, USA
[e] Department of Veterinary Surgery and Animal Reproduction, School of Veterinary Medicine and Animal Science, São Paulo State University (UNESP), São Paulo, SP 18618-681, Brazil

ARTICLE INFO

Keywords:
Facial expressions
Feline behavior
Feline welfare
Pain
Pain assessment
Scale dimensionality

ABSTRACT

The Feline Grimace Scale (FGS) is a facial expression-based scoring system for acute pain assessment in cats. This study aimed to investigate the dimensionality, importance of each action unit (AU), and variables affecting pain assessment using the FGS. One hundred images of cat faces were scored using the FGS by five veterinarians and five veterinary students. Cats were classified as painful or pain-free, whether the cut-off for analgesia was reached during real-time assessment. Scale dimensionality was studied using principal component analysis and Horn's parallel analysis. Item-total correlation investigated correlations between each AU and total FGS scores. Linear mixed models assessed responsiveness for each AU and variables influencing scores (age, gender, pain, and group). AUs had loading values $\geq$ 0.6, demonstrating an association for each AU with the first principal component of the PCA. All AUs and the FGS total ratio scores were increased in painful versus pain-free cats ($p < 0.001$). Female raters gave higher FGS scores than male raters ($p = 0.02$). Muzzle tension was the only AU with sensitivity below 70 %, whereas whiskers change was the only AU with specificity below 70 %. Similarly, whiskers change and muzzle tension had the lowest area under the curve values and Youden index. The FGS is a unidimensional scale, with total scores influenced by the rater's gender and pain. FGS demonstrated strong consistency and a high correlation between the AUs and total scores. However, muzzle tension and whiskers change are less discriminative than other AUs.

1. Introduction

The Feline Grimace Scale (FGS) is a valid and reliable instrument used for acute pain assessment based on changes in facial expressions in cats (Evangelista et al., 2020, 2019; Evangelista and Steagall, 2021; Watanabe et al., 2020). The tool consists of five action units (AU): ear position, orbital tightening, muzzle tension, whiskers change, and head position (scores 0–2 for each AU) (Evangelista et al., 2019). Construct validity was demonstrated by known-groups discrimination, and FGS scores were significantly higher in painful vs control cats. The FGS presented a very strong correlation with the Glasgow Composite Pain Scale – Feline. The intra-class correlation coefficient (ICC) was 0.89 (95 % CI: 0.85–0.92) for total scores when calculating inter-rater reliability and above 0.91 for intra-rater reliability. The instrument has an excellent discriminatory ability to differentiate between painful and non-painful cats, and a cut-off of > 0.39 out of 1 is suggested for the administration of analgesia (Evangelista et al., 2019). Sedation with dexmedetomidine-butorphanol, as well as anesthesia with propofol-isoflurane, influenced FGS scores 20 min after administration and up to 30 min after extubation, respectively (Watanabe et al., 2022). The instrument has shown adequate inter-rater reliability in cats undergoing dental extractions, and it is not affected by the presence of a

* Corresponding author at: Department of Clinical Sciences, Faculty of Veterinary Medicine, Université de Montréal, 3200 rue Sicotte, Saint-Hyacinthe, QC J2S 2M2, Canada.
E-mail address: pmortens@cityu.edu.hk (P.V. Steagall).

https://doi.org/10.1016/j.tvjl.2025.106448
Accepted 22 September 2025
Available online 23 September 2025

caregiver in the hospital setting (Watanabe et al., 2020). Inter-rater reliability was also good for FGS total scores (ICC > 0.80) when the instrument was used by veterinarians, veterinary students and nurses with very good agreement between these groups and veterinarians (Evangelista and Steagall, 2021). This finding was corroborated by a global survey when FGS total scores presented good reliability between cat caregivers and veterinarians (Monteiro et al., 2023). Construct validity, responsiveness, and reliability of the FGS were reported in kittens with ICC values for inter-rater reliability lower (0.68) than in previous studies in adults (Cheng et al., 2023). Finally, training can improve inter-rater reliability and agreement when the FGS is used by veterinarians, demonstrating the benefit of education in the subject (Robinson and Steagall, 2024).

Regardless of the extensive literature on the validity of the FGS for acute pain assessment in cats, the importance of each AU needs to be further studied. For example, the sensitivity and specificity of each AU as well as their discriminatory ability to differentiate painful versus non-painful individuals are not known. In a previous study, FGS total scores were significantly different before and after rescue analgesia showing responsiveness (Evangelista et al., 2019). However, construct validity including responsiveness of each AU (known-groups discrimination; painful versus pain-free) has not been determined.

Item-total correlation and internal consistency are related and important to assess the interrelationship of different items in a given instrument, but they assess distinct aspects of an instrument; the first demonstrates the correlation between each item and the total score, when the target item is excluded while the latter assesses overall consistency of items of a scale (Evangelista et al., 2019; Janse et al., 2021). The Cronbach's alpha coefficient for total FGS scores was 0.89 showing excellent internal consistency (Evangelista et al., 2019); this was minimally affected when removing one of the AU. Therefore, all AU contribute similarly to the FGS final score. However, item-total correlation has not been determined for assessing the correlation between each AU and the total FGS scores, and how well each AU discriminates painful vs non-painful cats.

Scale dimensional analysis provides a framework to understand scale homogeneity, its components and if different parts of a scale measure different attributes (Jollife and Cadima, 2016; Schatzman, 1991). Scale dimensional structure is important in pain assessment because pain is a complex and multidimensional experience, involving affective-emotional, cognitive and sensory aspects. A scale can assess a single (unidimensional) (e.g. intensity) or multiple (multidimensional) (e.g. intensity, location, impact on activity) aspects of a construct (e.g. pain) (Crawford Clark et al., 2002). The FGS likely assesses pain intensity, but a scale dimensional analysis has never been conducted. Understanding the dimensionality of the scale, the importance of each AU (i.e. discriminatory ability and responsiveness) and their correlations with total FGS scores, and how categorical variables affect assessment may add validity to the FGS in clinical practice.

The objectives of this study were to investigate scale dimensional structure and item-total correlation of the FGS. Additionally, the responsiveness and discriminatory ability of each AU considering specific categorical variables were determined. The hypotheses were that the FGS would be a unidimensional scale with high item-total correlation for each AU. Scores for each AU would be significantly different between painful and non-painful cats and before and after analgesia. Finally, pain category, gender, and age would affect FGS AU and total scores.

2. Materials and methods

2.1. Ethical approval

Data used in this study were collected from three previous studies after approval of the Institutional Animal Care and Use Committee of the Université de Montréal (protocols 17-Rech-1863; 17-Rech-1890; and 18-Rech-1825).

2.2. Images and FGS scores

One hundred face images of cats with different pain severity were included. These images had been selected from our databank and used during a previous study assessing the agreement and reliability of the FGS among cat owners, veterinarians, veterinary students, and nurses (Evangelista and Steagall, 2021). In the latter study, five raters from each group scored these 100 images using the FGS in an internet-based survey. For each image, each AU (ear position, orbital tightening, muzzle tension, whiskers change, and head position) was scored as 0 (AU absent), 1 (AU moderately present or uncertainty over its presence), or 2 (AU markedly present). If an AU was not clearly visible, raters could choose not to score that AU. For the present study, scores from veterinary students (5 female raters) and veterinarians (3 female; 2 male raters) were included as these two groups presented high inter-rater reliability for total FGS ratio scores: [owners—ICC single = 0.80 (95 % CI: 0.74–0.85); students—ICC single = 0.88 (95 % CI: 0.85–0.91); nurses—ICC single = 0.83 (95 % CI: 0.79–0.88); veterinarians—ICC single: 0.86 (95 % CI: 0.81–0.90)] (Evangelista and Steagall, 2021). The FGS total ratio score was obtained by the sum of all AUs from each rater divided by the number of AU that had been scored (Evangelista and Steagall, 2021).

2.3. Categorization of pain

Images were collected during three different prospective studies [Study A (Evangelista et al., 2019); Study B (Watanabe et al., 2020) and Study C (Evangelista et al., 2020)] involving pain-free (healthy controls or images collected before elective surgery) and painful cats (naturally occurring medical or surgical pain). During these studies, pain assessment was performed in real-time at several time-points (i.e. before and after surgery; before and after administration of rescue analgesia) using three instruments with reported validity for acute pain assessment in cats: the Glasgow Composite Measure Pain Scale-Feline (CMPS-F; Study A; female observer) (Calvo et al., 2014), the definitive CMPS-F (rCMPS-Feline; Study B; male observer) (Reid et al., 2017) and the short form of the Unesp-Botucatu Feline Pain Scale (UFEPS-SF) (Study C; female observer) (Belli et al., 2021). Immediately before or after pain assessment, cats were filmed undisturbed in their cages when the observer was absent. Screenshot images from these videos were obtained when cats were facing the camera (Evangelista et al., 2019). If real-time pain assessment indicated that a cat had reached the cut-off score (≥ 4/16 for CMPS-F; ≥ 5/20 for rCMPS-Feline; ≥ 4/12 for UFEPS-SF), analgesia was administered according to the study's protocol involving opioid and/or nonsteroidal anti-inflammatory drugs. For the present study, cats were categorized as being painful or pain-free according to whether they had or had not, respectively, reached the cut-off for administration of rescue analgesia. Real-time pain assessment allows the identification of changes in facial expressions, body posture, activity and demeanour of cats; observers are able to interact with cats and perform gentle palpation of wound/painful area and observe behavioural responses. Therefore, in this study, real-time pain scores by experienced observers using validated instruments was used for categorization of images as painful or pain-free.

2.4. Statistical analysis

Statistical analyses were performed using the integrated development environment RStudio (RStudio., 2025). Scores of FGS and its action unit are presented as means and standard deviations (mean ± SD), separately. Principal component analysis (PCA) and Horn's parallel analysis with 100 iterations were performed to assess the number of dimensions of the FGS (Dinno, 2009; Jollife and Cadima, 2016; Schatzman, 1991). In Horn's parallel analysis, dimensions were retained based on the comparison of eigenvalues from real data with those from simulated (random) data (Dinno, 2009; Preacher and MacCallum, 2003). Dimensions from the real data with an eigenvalue higher than the simulated data were retained (Goretzko and Ruscio, 2024; Preacher and MacCallum, 2003). The PCA yields loading values with coefficients ranging from −1–1. Thus, the closer to extremes, the more important the variable is. A cut-off of 0.6 was used (Hair et al., 2006) and loading values ≤ -0.60 and ≥ 0.60 were considered as significant association between the AU and the principal component or dimension (Hair et al., 2006). Item-total correlation was analysed using Spearman rank correlation coefficient for assessing the correlations between each AU (after excluding the target AU from the data) and the FGS total score within different pain conditions with the suitable values considered between 0.3 and 0.7 (Streiner et al., 2015). Construct validity was assessed using a multilevel linear mixed model for the FGS total scores and each AU. A linear mixed model was used to analyze the effects of categorical variables on the FGS total scores. Pain condition (painful vs. pain-free), group (veterinarians vs. student), age (21–29; 30–39; and 40–49 years) and rater gender (male vs. female) were considered fixed factors. The cat image was considered a random effect. Generalized linear mixed models were used to analyse the effects of these same variables on the scores of each AU separately. Similarly, pain condition, group, age, and gender of raters were considered fixed factors; cat image was considered a random factor. Adjustments for multiple comparisons used Bonferroni correction. A receiver operating characteristic (ROC) curve was calculated using Bootstrap with 1001 replications. Within this analysis, sensitivity, specificity, Youden index (refers to the concomitant highest values for sensitivity and specificity using the ROC) and area under the curve (AUC) were calculated for defining true positive and true negative scores (painful vs. pain-free) and the representation between true positive and false positive. The AUC ≥ 0.95 consisted of high discriminatory capacity (Streiner et al., 2015).

3. Results

A dataset containing 1000 observations (100 images each assessed by 10 different observers) was created. A total of 97 observations for which the observer had not given a score for one or more AU were excluded; 12 additional observations were excluded as they were considered outliers. These referred to images in which the evaluator had given FGS scores ≥ 8 but the real-time scores for that cat using the CMPS-F, rCMPS-Feline or UFEPS-SF did not indicate the presence of pain (i.e. scores were lower than the suggested cut-off score for administration of rescue analgesia for that scale). Therefore, 891 observations were included in the analysis. Mean FGS scores for painful and pain-free cats were 5.44 and 0.95 respectively (Table 1).

3.1. Scale dimensionality using principal component analysis

The PCA and Horn's parallel analysis showed that the FGS is unidimensional (Figs. 1 and 2). In Horn's parallel analysis, one dimension from the real data had an eigenvalue higher than the simulated data (3.6 vs 1.1). All other dimensions from the real data had eigenvalues lower than those of the simulated data as shown in Fig. 2. All AU had loading values ≥ 0.6, demonstrating a significant association for each AU with the first principal component (or first dimension) of the PCA (0.86, 0.83, 0.87, 0.83 and 0.84 for ear position, orbital tightening, muzzle tension, whiskers change and head position, respectively).

3.2. Item-total correlation

The item-total correlation coefficients (using Spearman rank correlation) were 0.69 when muzzle tension was excluded for pain-free cats, and 0.51 for painful cats. All coefficients fell within the range of 0.30–0.70, indicating that there are no inflated or redundant items in the scale (Table 2).

3.3. Construct validity for each AU and FGS total ratio scores and variables affecting scoring

All AU and the FGS total ratio scores were significantly increased in painful versus pain-free ($p < 0.001$). Female raters gave higher FGS scores than male raters ($p = 0.02$). Pain category (painful or pain-free) affected each individual AU. The AU ear position was not affected by any other variable. Orbital tightening ($p < 0.001$) and whiskers position ($p = 0.002$) were affected by gender; male observers gave lower scores than female observers (Table 3).

3.4. Specificity, sensitivity, and discriminatory ability of each AU and total scores

Table 4 presents sensitivity, specificity, Youden index, and AUC based on the ROC curve. Muzzle tension was the only AU with sensitivity below 70 %, whereas whiskers change was the only AU with specificity below 70 %. Similarly, whiskers change and muzzle tension had the lowest AUC values and Youden Index.

4. Discussion

This study investigated dimensionality structure, importance of each action unit (AU), and the variables affecting pain assessment using FGS. Scores of all AUs and the total FGS ratio scores were higher in painful than pain-free cats. These findings demonstrate the clinical applicability and responsiveness of the FGS but also of its individual AUs. These results are similar to previous studies on using FGS total ratio scores in cats undergoing ovariohysterectomy, both in real-time and with image scoring (Evangelista et al., 2020). However, it is the first time that the responsiveness and discriminatory ability of each AU is demonstrated. Item-to-item correlations showed that when one AU score is high, there is a general alignment with increased scores for other AUs, resulting in a consistent representation of the pain category. A cohesive set of AUs ensures that the assessment remains stable and valid across various contexts and observers; it demonstrates the importance and representative value of each AU to the full scale.

The PCA and Horn's parallel analysis confirm that the FGS is a unidimensional instrument. PCA focuses on the most significant features of

Table 1
Mean and standard deviations of total FGS scores and each action unit (AU) at variable category level.

Variable	Category	Total FGS Scores	Ear position	Orbital tightening	Muzzle tension	Whiskers change	Head position
Pain	Painful	5.44 ± 2.31	1.30 ± 0.64	1.23 ± 0.66	1.46 ± 0.53	1.28 ± 0.61	1.26 ± 0.62
	Pain free	0.95 ± 1.11	0.08 ± 0.27	0.13 ± 0.33	0.31 ± 0.48	0.24 ± 0.43	0.18 ± 0.39
Gender	Male	1.96 ± 2.43	0.45 ± 0.69	0.37 ± 0.66	0.51 ± 0.65	0.42 ± 0.63	0.50 ± 0.71
	Female	2.87 ± 2.79	0.57 ± 0.75	0.61 ± 0.73	0.81 ± 0.76	0.69 ± 0.72	0.62 ± 0.72
Age (years)	21–29	2.88 ± 2.78	0.57 ± 0.74	0.58 ± 0.73	0.82 ± 0.76	0.71 ± 0.73	0.63 ± 0.71
	30–39	2.59 ± 2.73	0.55 ± 0.76	0.55 ± 0.72	0.69 ± 0.72	0.56 ± 0.68	0.60 ± 0.75
	40–49	1.77 ± 2.54	0.40 ± 0.65	0.39 ± 0.65	0.46 ± 0.66	0.44 ± 0.67	0.38 ± 0.66
Group	Student	2.93 ± 2.81	0.56 ± 0.75	0.61 ± 0.74	0.82 ± 0.76	0.70 ± 0.73	0.64 ± 0.71
	Veterinarian	2.44–2.68	0.53–0.73	0.50–0.70	0.67–0.73	0.57–0.69	0.55–0.73

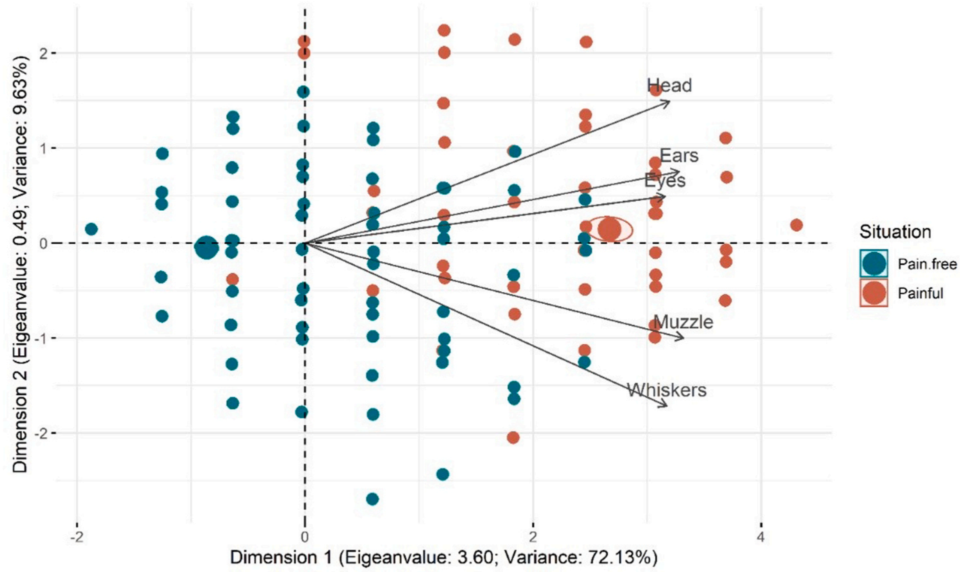

Fig. 1. Biplot of the principal component analysis for each action unit (AU) of the Feline Grimace Scale and the two categories (pain and pain-free). Each small point represents a cat image score. The direction of the vectors indicates the influence of each AU. Ellipses were produced to show the density of cats' image score for each situation (pain and pain-free). The green ellipse is small enough to fit entirely within the large green circle that crosses the negative part of the x axis circle and not easily visible. However, a slight inconsistency can be seen at the edge of the large green circle, which actually corresponds to the boundary of the green ellipse. Small circles indicate each individual assessment and large circles represent the centroid (center of mass given by the polygon formed with the interpolation of small circles of the same color).

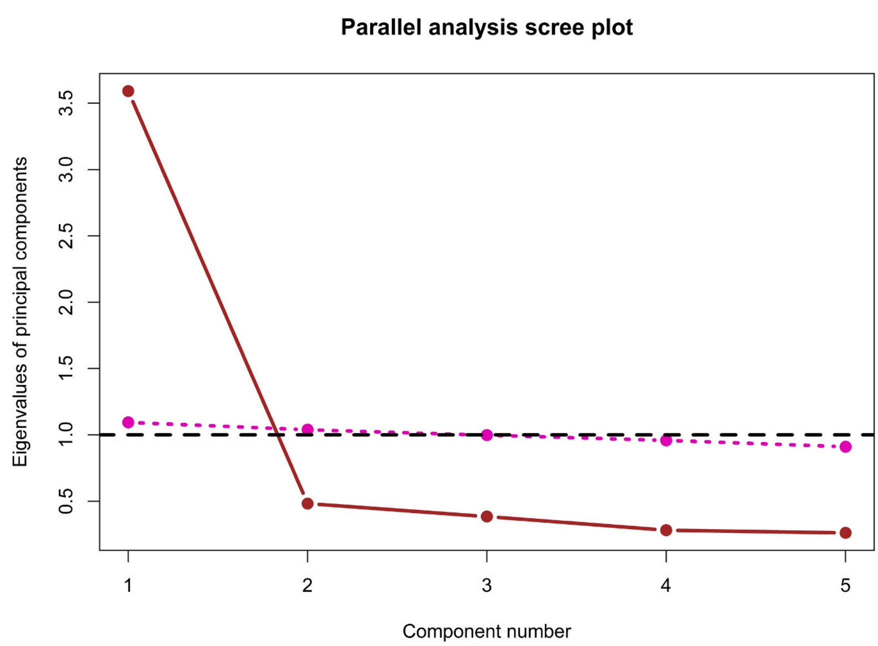

Fig. 2. Horn's parallel analysis is presented as scree plot. Brown line and dots show the actual data, whereas, pink dotted line show the simulated data. Brown and pink dots show the eigenvalues of the principal components of the actual and simulated data. Only one component of the actual data (brown dots) is above the components of the simulated data (pink dots).

data (Jollife and Cadima, 2016). In this study, each action unit (AU) exhibited high and similar loading values (0.83–0.87), indicating that each AU carries equal importance within the FGS. The FGS presented excellent internal consistency (0.89) when using the Cronbach's alpha coefficient (Evangelista et al., 2019). Moreover, alpha values were minimally affected by the removal of any single AU (Evangelista et al., 2019), corroborating that all AUs contribute equally to the final score. The strong associations of the AUs with PCA underscore their significance in accurately reflecting pain (Jollife and Cadima, 2016). Furthermore, the unidimensional nature of the scale simplifies the training process because of reduced complexity and simplified assessment procedure (Hagell, 2014).

High and positive Spearman rank coefficients among AUs and the total FGS scores demonstrated a strong overall correlation between each AU and the FGS total scores, and homogeneity of the scale. Values ranging from 0.3 to 0.7 showed that items had sufficient correlation ($>$ 0.37) without surpassing 0.70, which could indicate that AUs would just restate each other (Streiner, 2015). These correlations confirm that these

Table 2
Item-total Spearman rank correlation coefficients (rho) for each AU of the FGS for pain and pain-free cats, indicating the correlation between AU (by excluding the data of the target AU) and total scores.

Variable	Pain-Free	Painful
Excluding ear position	0.46	0.39
Excluding orbital tightening	0.47	0.37
Excluding muzzle tension	0.69	0.51
Excluding whiskers change	0.50	0.54
Excluding head position	0.48	0.41

AU are consistent during pain assessment in both painful and non-painful cats and this consistency continues even when one AU is excluded during pain assessment. The high correlation coefficient observed for all AUs supports findings from PCA, where loading values were similar across AUs demonstrating again that each AU plays an important role without redundancy; one AU is not more important than the other.

This study confirmed construct validity of each AU and their independent discriminatory capacity for pain categories. Previously, responsiveness as part of construct validity was demonstrated when FGS total scores were significantly different before and after rescue analgesia in kittens and adult cats (Cheng et al., 2023; Evangelista et al., 2020). Of interest, both AU and total ratio scores were not affected by group or age, except muzzle tension, demonstrating that discrimination of pain category using the AU and total ratio scores are important independently of experience (e.g. veterinarians or veterinary students) and perhaps years after graduation, as veterinarians tend to be older than veterinary students. Age or group were also not important for ear and head position scoring, and indeed, these are two AU that have demonstrated the highest reliability across different groups of individuals (Evangelista and Steagall, 2021).

Female raters assigned higher FGS scores than male raters. As much as these results merit discussion, one should bear in mind that the present study was designed to investigate gender differences and acute pain scores. A total of eight female and two male individuals participated in our study; there were five female veterinary students, and the two male individuals were both from the group of veterinarians. The unequal ratio and small sample may have biased these results. These issues may also be applied to orbital tightening and whiskers change scoring that were affected by gender. Previous studies have suggested that pain assessment in cats and dogs may be influenced by the rater's gender (Christov-Moore et al., 2014; Evangelista and Steagall, 2021; Sadeghiyeh et al., 2012). In particular, FGS scores were significantly higher among female raters (comprising cat owners, veterinarians, veterinary students, and nurses) compared to their male counterparts (Evangelista and Steagall, 2021). Extensive literature exists on gender differences in responses to pain in humans (Sadeghiyeh et al., 2012). Research indicates that females are generally more empathetic toward pain and distress than males (Christov-Moore et al., 2014; Sadeghiyeh et al., 2012). Moreover, female veterinarians are more likely to provide higher pain scores and administer analgesics for dogs and cats compared to males (Beswick et al., 2016; Evangelista and Steagall, 2021; Mich et al., 2010). These findings may provide insights into the influence of gender on feline pain assessment, warranting further exploration. However, a large bilingual survey of cat caregivers found no gender effect on pain assessment when using the FGS (Monteiro et al., 2023). The absence of a difference in scores between male and female participants in that study may suggest that cat caregivers perceive pain similarly, irrespective of gender. Our laboratory is currently performing a study investigating the effects of gender on acute pain assessment in cats involving large and balanced sample sizes.

The muzzle and whiskers are less discriminative, exhibiting lower sensitivity and specificity, respectively, compared to other AUs. Furthermore, whiskers change and muzzle tension had the lowest AUC

Table 3
Results of linear mixed model using FGS total scores and generalized linear mixed models using scores from each action unit (ear position, orbital tightening, muzzle tension, whiskers change, and head position). Categories 'pain-free', 'female', '21–29 years age', and 'student' were reference categories in each model.

Variables	Categories	Total FGS score		Ear position		Orbital tightening		Muzzle tension		Whiskers change		Head position	
		Estimates	p values	Estimates	p values	Estimates	p values	Estimates	p values	Estimates	p values	Estimates	p values
Pain	Painful	3.14	< 0.001	0.62	< 0.001	0.43	< 0.001	0.88	< 0.001	0.74	< 0.001	0.38	< 0.001
Gender	Male	−0.39	0.02	−0.008	0.81	−0.23	< 0.001	−0.08	0.13	−0.17	0.002	0.03	0.47
Age (years)	30–39	0.19	0.27	−0.005	0.90	0.22	< 0.001	−0.04	0.43	−0.09	0.11	0.09	0.03
	40–49	−0.10	0.71	−0.05	0.38	0.28	< 0.001	−0.14	0.09	−0.04	0.65	−0.07	0.24
Group	Veterinarian	−0.19	0.22	0.03	0.34	−0.16	< 0.001	−0.003	0.94	0.07	0.17	−0.12	0.002

Table 4
Sensitivity, specificity, area under the curve (AUC), and Youden Index for each action unit. Data are presented as estimates and 95 % confidence intervals.

	Sensitivity	Specificity	Area under the curve	Youden Index
Ear position	0.98 (0.95–1.0)	0.80 (0.77–0.83)	0.94 (0.93–0.96)	0.78 (0.73–0.83)
Orbital tightening	0.92 (0.88–0.95)	0.75 (0.72–0.79)	0.88 (0.86–0.91)	0.67 (0.61–0.74)
Whiskers change	0.93 (0.90–0.96)	0.65 (0.61–0.69)	0.87 (0.84–0.89)	0.59 (0.52–0.65)
Muzzle tension	0.63 (0.57–0.72)	0.95 (0.93–0.97)	0.87 (0.85–0.90)	0.59 (0.50–0.68)
Head position	0.94 (0.91–0.98)	0.71 (0.68–0.74)	0.89 (0.87–0.91)	0.66 (0.59–0.72)
FGS total ratio scores	0.92 (0.88–0.96)	0.92 (0.89–0.94)	0.97 (0.97–0.98)	0.84 (0.77–0.90)

and Youden Index. Muzzle and whiskers assessment can be challenging when using images, as factors such as camera position, coat color, lighting, image resolution, and pixels can affect the evaluation. Low-quality images may have a greater impact on scoring these AUs, than on more prominent features like head position, ear position, and orbital tightening. Previous studies have shown that the reliability of scores for the muzzle and whiskers is lower than that of other AUs, with ICCs of 0.56 for muzzle position and 0.64 for whiskers change in cats (Watanabe et al., 2020). In kittens, the ICCs were even lower than in adult cats (0.47 for muzzle position and 0.35 for whiskers change) (Cheng et al., 2023). Similarly, inter-rater reliability was also lower in a large bilingual survey of cat caregivers (Monteiro et al., 2023). In that study, the mean scores for each AU and the total FGS scores were not significantly different between cat caregivers and veterinarians, with the exception of muzzle tension (Monteiro et al., 2023). This discrepancy may be attributed to the difficulty cat caregivers faced in scoring muzzle tension. In a previous survey, whiskers change had the highest percentage of "not possible to score" responses, accounting for 10.2 % of the images, while muzzle tension had the second highest prevalence of 3.6 % (Evangelista et al., 2019). This may be because image assessment lacks dimensional perspective, which would aid in accurately identifying the positions of the whiskers and muzzle. These issues may be circumvented by continuing education as training on FGS scoring have improved inter-rater reliability for both muzzle tension and whiskers change (Robinson and Steagall, 2024). On the other hand, ear position, orbital tightening and head position demonstrated sensitivity and specificity above 0.70, which may better discriminate true positives and negatives. The AUC for FGS total scores were > 0.95 demonstrating again an excellent discriminatory ability of the instrument for feline acute pain assessment.

5. Conclusion

The FGS is a unidimensional scale. Construct validity was demonstrated for each AU and total FGS scores; the FGS AUs have strong consistency, demonstrating a high correlation with the total FGS ratio scores. These findings emphasize the significance of each AU without redundancy. Pain scoring was affected by rater's gender, but this should be interpreted with caution giving the small sample size and unequal group ratios. Muzzle position may be less discriminative of true positives, whereas whiskers change may affect true negatives. However, all AUC demonstrated that AUs and especially total ratio scores have excellent discriminatory ability to differentiate painful and non-painful cats. Our current research is investigating how deep learning models can reduce bias related to AU scoring, and gender, experience and age of raters, while eliminating the need for scoring some AUs when using poor quality images (Steagall et al., 2023).

CRediT authorship contribution statement

Luna Stelio PL: Writing – review & editing, Visualization, Investigation, Conceptualization. **Syed S.U.H. Bukhari:** Writing – review & editing, Writing – original draft, Methodology, Data curation. **Trindade Pedro H:** Writing – review & editing, Resources, Methodology, Investigation. **Monteiro Beatriz P:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Steagall Paulo V:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization.

Ethics approval

This study was approved by the Institutional Animal Care and Use Committee of the Université de Montréal (protocols 17-Rech-1863; 17-Rech-1890; and 18-Rech-1825).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

None.

Data availability

The data that support the findings of this study are available on request from the corresponding author.

References

Belli, M., de Oliveira, A.R., de Lima, M.T., Trindade, P.H.E., Steagall, P.V., Luna, S.P.L., 2021. Clinical validation of the short and long UNESP-Botucatu scales for feline pain assessment. PeerJ 9, 1–21. https://doi.org/10.7717/peerj.11225.

Beswick, A., Dewey, C., Johnson, R., Dowsett-Cooper, J., Niel, L., 2016. Survey of ontario veterinarians' knowledge and attitudes on pain in dogs and cats in 2012. Canadian Veterinary Journal 57, 1274.

Calvo, G., Holden, E., Reid, J., Scott, E.M., Firth, A., Bell, A., Robertson, S., Nolan, A.M., 2014. Development of a behaviour-based measurement tool with defined intervention level for assessing acute pain in cats. Journal of Small Animal Practice 55, 622–629. https://doi.org/10.1111/jsap.12280.

Cheng, A.J., Malo, A., Garbin, M., Monteiro, B.P., Steagall, P.V., 2023. Construct validity, responsiveness and reliability of the feline grimace scale in kittens. Journal of Feline Medicine and Surgery 25. https://doi.org/10.1177/1098612X231211765.

Christov-Moore, L., Simpson, E.A., Coudé, G., Grigaityte, K., Iacoboni, M., Ferrari, P.F., 2014. Empathy: gender effects in brain and behavior. Neuroscience & Biobehavioral Reviews 46, 604–627. https://doi.org/10.1016/j.neubiorev.2014.09.001.

Crawford Clark, W., Yang, J.C., Tsui, S.L., Ng, K.F., Bennett Clark, S., 2002. Unidimensional pain rating scales: a multidimensional affect and pain survey (MAPS) analysis of what they really measure. Pain 98, 241–247. https://doi.org/10.1016/S0304-3959(01)00474-2.

Dinno, A., 2009. Exploring the sensitivity of Horn's parallel analysis to the distributional form of random data. Multivariate Behavioral Research 44, 362–388. https://doi.org/10.1080/00273170902938969.

Evangelista, M.C., Benito, J., Monteiro, B.P., Watanabe, R., Doodnaught, G.M., Pang, D.S.J., Steagall, P.V., 2020. Clinical applicability of the feline grimace scale: Real-time versus image scoring and the influence of sedation and surgery, 2020 PeerJ 1–17. https://doi.org/10.7717/peerj.8967.

Evangelista, M.C., Steagall, P.V., 2021. Agreement and reliability of the feline grimace scale among cat owners, veterinarians, veterinary students and nurses. Scientific Reports 11, 1–9. https://doi.org/10.1038/s41598-021-84696-7.

Evangelista, M.C., Watanabe, R., Leung, V.S.Y., Monteiro, B.P., O'Toole, E., Pang, D.S.J., Steagall, P.V., 2019. Facial expressions of pain in cats: the development and validation of a feline grimace scale. Scientific Reports 9, 1–11. https://doi.org/10.1038/s41598-019-55693-8.

Goretzko, D., Ruscio, J., 2024. The comparison data forest: a new comparison data approach to determine the number of factors in exploratory factor analysis. Behavior Research Methods 56, 1838–1851. https://doi.org/10.3758/s13428-023-02122-4.

Hagell, P., 2014. Testing rating scale unidimensionality using the principal component analysis (PCA)/<I>t</I>-Test protocol with the rasch model: the primacy of theory over statistics. Open Journal of Statistics 04, 456–465. https://doi.org/10.4236/ojs.2014.46044.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L., et al., 2006. Multivariate Data Analysis, sixth ed. Pearson Prentice Hall, New Jersey.

Janse, R.J., Hoekstra, T., Jager, K.J., Zoccali, C., Tripepi, G., Dekker, F.W., Van Diepen, M., 2021. Conducting correlation analysis: important limitations and pitfalls. Clinical Kidney Journal 14, 2332–2337. https://doi.org/10.1093/ckj/sfab085.

Jollife, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374. https://doi.org/10.1098/rsta.2015.0202.

Mich, P.M., Hellyer, P.W., Kogan, L., Schoenfeld-Tacher, R., 2010. Effects of a pilot training program on veterinary students' pain knowledge, attitude, and assessment skills. Journal of Veterinary Medical Education 37, 358–368. https://doi.org/10.3138/jvme.37.4.358.

Monteiro, B.P., Lee, N.H.Y., Steagall, P.V., 2023. Can cat caregivers reliably assess acute pain in cats using the feline grimace scale? A large bilingual global survey. The Journal of Feline Medicine and Surgery 25. https://doi.org/10.1177/1098612X221145499.

Preacher, K.J., MacCallum, R.C., 2003. Repairing tom Swift's electric factor analysis machine. Understanding Statistics Stat Issues Psychol Educ Soc Sci, pp. 13–43 vol. 2.

Reid, J., Scott, E.M., Calvo, G., Nolan, A.M., 2017. Definitive glasgow acute pain scale for cats: validation and intervention level. Veterinary Record 180, 449. https://doi.org/10.1136/vr.104208.

Robinson, A.R., Steagall, P.V., 2024. Effects of training on feline grimace scale scoring for acute pain assessment in cats. Journal of Feline Medicine and Surgery 26, 1–7. https://doi.org/10.1177/1098612X241275284.

RStudio. [WWW Document], 2025. URL ⟨https://www.rstudio.com/⟩ (Accessed 1.1.25).

Sadeghiyeh, H., Khorrami, A., Hatami, J., 2012. Gender differences in empathy for pain: a single-pulse transcranial magnetic stimulation study. Clinical Neurophysiology 42, 62. https://doi.org/10.1016/j.neucli.2011.11.029.

Schatzman, L., 1991. Dimensional analysis: notes on an alternative approach to the grounding of theory in qualitative research. In: Social Organization and Social Process. Routledge, Chicago, USA, pp. 303–314.

Steagall, P.V., Monteiro, B.P., Marangoni, S., Moussa, M., Sautié, M., 2023. Fully automated deep learning models with smartphone applicability for prediction of pain using the feline grimace scale. Scientific Reports 13, 1–13. https://doi.org/10.1038/s41598-023-49031-2.

Streiner, D., Norman, G., Cairney, J., 2015. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford University Press Inc, New York. https://doi.org/10.1093/acprof:oso/9780199231881.003.0006.

Streiner, D.L., 2015. Best (but oft-forgotten) practices: the multiple problems of multiplicity-whether and how to correct for many statistical tests. The American Journal of Clinical Nutrition 102, 721–728. https://doi.org/10.3945/ajcn.115.113548.

Watanabe, R., Doodnaught, G.M., Evangelista, M.C., Monteiro, B.P., Ruel, H.L.M., Steagall, P.V., 2020. Inter-Rater reliability of the feline grimace scale in cats undergoing dental extractions. Frontiers in Veterinary Science 7, 4–9. https://doi.org/10.3389/fvets.2020.00302.

Watanabe, R., Monteiro, B.P., Ruel, H.L.M., Cheng, A., Marangoni, S., Steagall, P.V., 2022. The effects of sedation with Dexmedetomidine–Butorphanol and anesthesia with Propofol–Isoflurane on feline grimace Scale© scores. Animals 12. https://doi.org/10.3390/ani12212914.