

RVC OPEN ACCESS REPOSITORY – COPYRIGHT NOTICE

This is the peer-reviewed, manuscript version of an article published in *Veterinary Record*. The final version is available online via <http://dx.doi.org/10.1136/vr.104245>.

The full details of the published version of the article are as follows:

TITLE: Veterinary student competence in equine lameness recognition and assessment: a mixed methods study

AUTHORS: Sandra D Starke, Stephan A May

JOURNAL TITLE: Veterinary Record

PUBLISHER: BMJ Publishing Group

PUBLICATION DATE: 12 August 2017 (online)

DOI: 10.1136/vr.104245

1 **Veterinary student competence in equine lameness recognition and assessment: a mixed**
2 **methods study**

3 Sandra D. Starke^{1,*}, MSc PhD and Stephen A. May¹, MA VetMB PhD DVR DEO FRCVS
4 DipECVS FHEA

5

6 ¹ The Royal Veterinary College, Hawkshead Lane, North Mymms, Hatfield, Hertfordshire,
7 AL9 7TA, UK

8

9 * Corresponding author (S.D. Starke): sstarke@rvc.ac.uk

10

11

12 **Abstract**

13 The development of perceptual skills is an important aspect of veterinary education. In this
14 study, we investigated veterinary student competency in lameness evaluation at two stages,
15 prior to (3rd year) and during (4th/5th year) clinical rotations. Students evaluated horses in
16 videos, where horses were presented on a straight line and on circles. Eye tracking data were
17 recorded to follow student gaze. On completing the task, students filled in a structured
18 questionnaire. Results showed that experienced veterinary students outperformed
19 inexperienced students in classifying the correct limb as lame, although performance of both
20 cohorts was inadequate in most scenarios. Mistakes largely arose from classifying an incorrect
21 limb as lame. The correct classification of sound horses as non-lame was at chance level.
22 While the experienced student cohort primarily looked at upper body movement (head and
23 sacrum) during lameness assessment, the inexperienced cohort focused on limb movement.
24 Student self-assessment of performance was realistic, and task difficulty was most commonly
25 rated between 3 and 4 out of 5. The inexperienced students named a considerably greater
26 number of visual lameness features than the experienced students. Future dedicated training
27 based on the findings presented here may help students to develop more reliable lameness
28 assessment skills.

29

30 **Introduction**

31 An important aspect of veterinary education is the development of perceptual skills. In equine
32 veterinary education, the ability to correctly identify, differentiate and classify lameness is
33 one such skill. However, even experienced veterinarians struggle with this task, especially for
34 mild lameness (Dow and others 1991; Keegan and others 1998; Ishihara and others 2005;
35 Orito and others 2007; Keegan and others 2010; Mc Cracken and others 2012). Currently,
36 only a small number of studies have investigated student performance in equine gait scoring
37 (Keegan and others 1998; Arkell and others 2006) and the nature of their visual assessments
38 remains largely unexplored. In order to develop a systematic approach to lameness education,
39 leading to reliable performance of our future practitioners, it would be beneficial to
40 understand student performance better.

41 The experience-dependent ability to assess gait abnormalities has been investigated in
42 numerous studies, covering not only horses (Keegan and others 1998; Arkell and others 2006)
43 but also humans (Read and others 2003; Brunnekreef and others 2005; Maathuis and others
44 2005; Brown and others 2008; Ong and others 2008; Williams and others 2009), dogs (Kestin
45 and others 1992) and pigs (Main and others 2000). Gait assessment skills improve with
46 experience. Intra-observer agreement (the same person judging a scenario multiple times)
47 often becomes more consistent (Keegan and others 1998; Main and others 2000; Brunnekreef
48 and others 2005; Arkell and others 2006; Ong and others 2008), although this may simply
49 reflect mistakes being made more consistently (Kawamura and others 2007; Waxman and
50 others 2008). Inter-observer agreement (different people judging the same scenario) also often
51 increases with experience (Main and others 2000; Brunnekreef and others 2005; Maathuis and
52 others 2005; Arkell and others 2006; Ong and others 2008). For visual gait assessment in
53 horses, a past landmark study highlighted the strong subjectivity and unreliability of gait
54 evaluation independent of experience level. The authors hypothesised that high intra-observer
55 variation in less experienced observers may be rooted in a lack of confidence and a less

56 systematic assessment protocol compared to their experienced counterparts (Keegan and
57 others 1998).

58 For a veterinary student, lameness assessment skills are commonly developed through a
59 combination of taught classes and practical exposure on clinical rotations and work
60 placements. During practical sessions, students learn from the experts of their respective
61 institutions. Dedicated textbooks support the expansion of knowledge regarding the gait
62 examination (Wyn-Jones 1988; Stashak and Hill 1996; Stashak 2002; Ross and Dyson 2011),
63 although authors vary in the weighting of different lameness pointers. At the Royal Veterinary
64 College (RVC), at the time of performing this study, the didactic aspects of equine lameness
65 were first covered in year 3 of the BVetMed curriculum. At that point, clinical observation
66 and decision making skills are starting to be developed. The gained knowledge is then applied
67 during intramural clinical rotations in years 4 and 5 where students spend one week on equine
68 orthopaedics. Here, students follow the clinician on duty when performing lameness
69 examinations of horses admitted to the RVC's Equine Referral Hospital (ERH). Students with
70 an interest in equine work typically also choose to perform their extramural studies (EMS) in
71 equine practice or referral hospitals, where a larger number of lameness cases can be
72 followed.

73 To date, gait assessment studies have focused largely on performance metrics as outcome
74 variables. However, during perceptual tasks it is equally important to quantify the approach
75 taken by participants in order to relate performance and approach. In the case of lameness
76 assessment, this approach would be quantifiable through the pattern of visual attention
77 allocation. Eye tracking allows quantification of where a participant is looking when making
78 judgements, opening the door to insights into cognitive processes that were previously
79 inaccessible. Due to the narrow field of sharp vision (1° to 2° visual angle for foveal,
80 respectively central, vision), humans have to look at those parts in a scene where they expect
81 to find the information that they are seeking (Findlay and Gilchrist 2003). Eye tracking has

82 been used in thousands of studies (Tien and others). In human medicine, for example, it has
83 helped to understand errors made when interpreting diagnostic images (Thomas and
84 Lansdown 1963; Kundel and others 1978; Krupinski 1996; Krupinski 2005; Krupinski 2010;
85 Balslev and others 2012).

86 In the present study, we investigated the approaches and abilities of a random sub-section of
87 veterinary students from two different stages in their degree programme when evaluating
88 horses for lameness. We used a combination of gait scoring, eye tracking and questionnaires
89 to quantify the percentage of horses evaluated correctly, sources of judgement error, visual
90 assessment strategies and student attitudes. The aim of this work is to provide a
91 comprehensive baseline for the development of more effective courses on equine lameness
92 identification.

93

94 **Materials and Methods**

95 ***Task***

96 The task was to assess horses on a PC monitor. Thirty-four video recordings (25 Hz,
97 resolution 1024 x 768 pixels) were selected to compile a database comprising an
98 approximately equal presence of fore- and hind limb lameness in both the left and right limbs
99 in different views (Figure 1): 15 clips showed the horses trotting away from and towards the
100 observer on a straight line (condition ‘straight line’), 15 horses trotted on a circle on either the
101 left or the right rein (condition ‘circle’) and four horses were shown trotting in lateral view
102 (condition ‘lateral’). The lateral condition was only used to explore student opinion regarding
103 this view, but not for in-depth analysis of performance. All video recordings were edited in
104 Premiere CS4 (Adobe, USA) into clips of approx. 30 to 50 s duration. Each clip presented the

105 horse in the chosen view without sound, the video being repeated at least once. Clips were
106 presented on a 21 inch CRT monitor (Dell Trinitron, refresh rate: 100 Hz).

107

108 **[Figure 1]**

109

110 ***Lameness***

111 The veterinary reports for all horses presented on straight line and circle were obtained in
112 order to establish the lame limb and lameness score of each horse from the expert clinical
113 examination based on visual examination and final diagnosis. Four video recordings of horses
114 filmed in lateral view were scored by three experienced assessors who agreed on the lame leg
115 independently and an average was taken from all three scores. For all horses classed as lame,
116 a visually detectable lameness in the limb consistent with the expert opinion was verified by
117 the experimenter (S D Starke) to warrant that it could be shown to students and ensure that
118 lameness was visible to the trained eye. Fourteen out of 15 horses were deemed lame on the
119 straight line, twelve out of 15 on the circle and three out of four in lateral view. Horses were
120 unilaterally fore- or hind limb lame with a lameness grade of up to 5 out of 10 (UK grading
121 scale, see Arkell and others or popular textbooks) across limbs and conditions. The lame
122 horses on the straight line had a mean (SD) lameness score of 2.38 (1.30) out of 10 for
123 forelimb lameness and 3.00 (1.26) out of 10 for hind limb lameness. On the circle this
124 amounted to 3.67 (1.21) out of 10 for forelimb lameness and 3.50 (1.05) out of 10 for hind
125 limb lameness and in lateral view to 3.67 (0.58) out of 10 for hind limb lameness.

126 ***Participants***

127 Twenty 3rd year BVetMed students (the ‘inexperienced’ cohort) as well as 12 4th/5th year
128 BVetMed students (the ‘experienced’ cohort, who had performed equine rotations and some

129 participants had done at least one EMS placement in an equine hospital or practice) were
130 recruited from the RVC student body to participate in the study. All participants volunteered
131 to take part in the study by responding to an email invitation (inexperienced and part of the
132 experienced cohort) or after additional encouragement from lecturing staff (part of the
133 experienced cohort).

134 ***Eye tracking and gait scoring***

135 Eye tracking data were collected in parallel to the scoring session using a stand-alone eye
136 tracker (X120, Tobii Technology, Sweden). The eye tracker was positioned underneath the
137 monitor and calibrated for each individual participant at the beginning of the assessment
138 session using the default 5-point calibration procedure. Eye tracking data were recorded at 60
139 Hz, with a small subset of data recorded at 120 Hz at the beginning of the study. The viewing
140 distance between student and eye tracker was approximately 70 cm, ranging dynamically
141 between 50 to 80 cm according to the calibration volume of the eye tracker.

142 Video clips were randomised in Matlab (The MathWorks, USA) and presented as seven
143 segments within the proprietary eye tracker software (Tobii Studio v.2). Each student was
144 presented with the seven segments, each containing several video clips, in random order. This
145 provided the best trade-off at the time between software limitations and workload arising from
146 randomisation. For the assessment of each horse, an on-screen questionnaire was shown after
147 each video clip with the following questions and single choice answers:

148 1. Is the horse lame? If it is lame, which is the most affected leg?

149 [Answers: Sound, Left fore, Right fore, Left hind, Right hind]

150 2. Which lameness score would you assign to that leg?

151 [Answers: 0/10 (Sound), 1/10, 2/10, 3/10, 4/10, 5/10, 6/10, 7/10, 8/10, 9/10,
152 10/10 (non weight bearing)]

153 3. Is a further leg affected?

154 [Answers: No, Left fore, Right fore, Left hind, Right hind]

155 ***Questionnaires***

156 After completion of the task, students were given a questionnaire which contained questions
157 relating to their self-evaluation and approaches when assessing horses for lameness. Students
158 were asked to:

- 159 • Rate the difficulty of the lameness assessment task on a scale from 1 (very simple) to
160 5 (very difficult).
- 161 • Estimate the percentage of horses they expected to have evaluated correctly.
- 162 • Rank the difficulty of assessing horses on the straight line, on a circle and in lateral
163 view.
- 164 • Note down what movement features were used when assessing forelimb and hind limb
165 lameness on the straight and circle.

166 The cohort of experienced students was additionally asked to estimate how many lameness
167 cases they had seen so far.

168 ***Analysis of gait scoring***

169 Scores for straight line and circle were analysed using custom-written software in Matlab
170 (The MathWorks, USA) after exporting data from Tobii Studio. For each student, the
171 response for each horse was compared to the gold standard derived from clinical records. The
172 following metrics were calculated to quantify performance: 1) the percentage of horses
173 correctly classified as lame and sound, 2) the percentage of horses classified lame in the
174 correct limb and 3) the percentage of incorrect classifications. Incorrect classifications were
175 attributed to the following five reasons: students considering a lame horse sound; students

176 considering a sound horse lame; and students selecting the incorrect limb(s) as affected,
177 further split into students selecting the incorrect contralateral limb as affected and students
178 selecting an incorrect ipsilateral limb as affected. Variation in the allocated grades was
179 quantified for each horse across all those students who had deemed the horse primarily lame
180 in the fore- or hind limb. Score range as well as the difference between each score and the
181 median score were calculated for each cohort. Agreement across students was calculated
182 using free marginal kappa (κ). Inter-observer agreement was calculated using the free
183 marginal multilayer kappa (Brennan and Prediger 1981; Randolph 2005; Warrens 2010) using
184 an online kappa calculator (<http://justusrandolph.net/kappa>). Kappa was calculated for the
185 presence or absence of lameness in general, forelimb lameness and hind limb lameness. For
186 this metric, the side of lameness was not considered.

187 Statistical analysis was performed in IBM SPSS 20 (IBM Corp.), where a Shapiro-Wilk test
188 was run to decide on the use of parametric or non-parametric tests. The percentages of lame
189 horses correctly classed as sound or lame and correctly classed as lame in the correct limb
190 was compared between student cohorts for straight and circle using the Mann-Whitney U
191 Test. Consistency in allocated grades was tested this way, too. Within student cohorts, a
192 Wilcoxon signed-rank matched-pairs test was used to examine differences between
193 performance for fore- and hind limb lameness and between straight line and circle. Regression
194 analysis was performed to determine whether there was an association between the caseload
195 seen and performance of the experienced cohort.

196 *Analysis of visual assessment*

197 For the assessment of the 15 horses trotting on a straight line away from / towards the
198 observer, eye tracking data were analysed to determine the preferred regions students examine
199 visually. The constant change in direction of horses on the circle proved prohibitive for
200 analysis due to the manual workload. In order to assign the gaze data to discrete regions on

201 the horse, the position of the os sacrum, head and feet in the video recordings were digitised
202 using Tracker (www.cabrillo.edu/~dbrown/tracker). In Matlab (The MathWorks, USA), point
203 of gaze was then calculated for each frame and assigned to an equally spaced horizontal grid,
204 where it was assigned to one out of 18 bins (Figure 2). The grid was calculated based on the
205 tracked os sacrum (trot away from observer) or head (trot towards observer) position with the
206 feet as reference points. For each participant and each of the 15 horses, the percentage
207 viewing time for each region was then calculated as the sum of frames spent looking at each
208 region divided by the total viewing time. The total viewing time was calculated as the sum of
209 all frames where gaze data were tracked on the monitor. For each participant, the mean
210 viewing time was then calculated for each of the 18 discrete regions across all videos in which
211 at least 60 % of the total video playing time was being tracked.

212 Statistical analysis was carried out in SPSS as described above. A Mann-Whitney U Test was
213 performed to compare the viewing times allocated to each of the 18 regions (Figure 2 and 5)
214 between the two student cohorts, separately for trot away from and towards the observer.

215

216 **[Figure 2]**

217

218 *Analysis of questionnaire responses*

219 From the answers to the questionnaires, the difficulty ratings for the task were extracted and
220 average values and frequencies calculated. The percentage of horses students thought they had
221 evaluated correctly (self-rating) and the objective percentage of horses evaluated correctly as
222 described above was calculated across all 34 horses and all three views. The difference
223 between self-assessment and objective assessment was calculated for each student and mean
224 differences for each scenario and student cohort were calculated to examine bias in self-

225 rating. The lameness pointers noted down by the students were extracted from the
226 questionnaires, counted and ranked by nomination for assessment on the straight line and
227 circle.

228 Statistical analysis was carried out in SPSS as described above. Differences in self-rating
229 were compared between the inexperienced and experienced student cohort using an
230 independent samples t-test.

231

232 **Results**

233 In this study we report results as the median [interquartile range, IQR] unless stated otherwise.
234 For those readers unfamiliar with these metrics, they are used for data that does not follow a
235 normal distribution (being non-parametric). The median reports the central tendency and is the
236 point found in the middle of an ordered distribution: it is the value which separates the 50% of
237 data points which are bigger from the 50% that are smaller. It is also called 2nd quartile: when
238 splitting the data distribution into four equal parts from lowest to highest value, quartiles mark
239 the thresholds between each of these four parts. The IQR reports data variation and quantifies
240 the spread between the 1st and 3rd quartile: it is the range in which the ‘middle’ 50% of data
241 points can be found, and is in meaning similar to the standard deviation in normal distributed
242 data. Further details can be found in standard statistics textbooks.

243 ***Gait scoring***

244 The median [IQR] percentage of lame horses correctly classed as generally lame by the
245 inexperienced and experienced cohort for straight line and circle combined was 92 [10] % and
246 98 [10] %, respectively. There was no significant difference between cohorts ($P \geq 0.307$). The
247 median [IQR] of sound horses correctly classed as sound across all views was 50 [30] % for
248 both cohorts, not being significantly different ($P = 0.924$).

249 For forelimb lameness, the median [IQR] of horses classified lame in the correct limb (Figure
250 3) for the inexperienced and experienced cohort was 50 [38] % and 69 [25] %, respectively,
251 on the straight as well as 83 [42] % and 100 [17] %, respectively, on the circle. For hind limb
252 lameness, these values decreased to 33 [33] % and 67 [25] %, respectively, on the straight and
253 50 [33] % and 67 [33] %, respectively, on the circle. Differences in percentage correct
254 between inexperienced and experienced cohort were significant ($P \leq 0.032$) except for
255 forelimb lameness on the circle ($P = 0.053$). Within student cohorts, differences were
256 significant between fore- and hind limb lameness on the circle ($P \leq 0.032$), but not on the
257 straight ($P \geq 0.313$) and between trotting on the straight and circle for forelimb lameness ($P \leq$
258 0.003), but not for hind limb lameness ($P \geq 0.253$).

259

260 **[Figure 3]**

261

262 Incorrect classifications were mainly the result of declaring an incorrect limb as lame, with a
263 median [IQR] of 71 [9] % and 60 [17] % for the inexperienced and experienced cohort,
264 respectively. The incorrect contralateral and ipsilateral limbs were selected in similar
265 proportion.

266 Mean grades did not differ significantly between cohorts for forelimb lameness ($P = 0.067$).

267 However, for hind limb lameness grade differed significantly between cohorts ($P < 0.001$),

268 where the experienced cohort's grade had a median score of one below the inexperienced

269 cohort. The inexperienced cohort did not grade more than six grades away from the group's

270 median grade, while the experienced student group graded no more than 3.5 grades away

271 (Figure 4).

272

273 **[Figure 4]**

274

275 Within each student cohort, agreement (Table 1) was highest for the presence or absence of
276 lameness on the circle ($\kappa = 0.70$ for inexperienced and $\kappa = 0.89$ for experienced students) and
277 lowest for the presence or absence of hind limb lameness during evaluation away from /
278 towards the observer ($\kappa = 0.15$ for inexperienced and $\kappa = 0.20$ for experienced students).

279 For the experienced student cohort, there was no association between the number of cases
280 seen as part of their degree and the percentage of correctly evaluated horses for the nine
281 students that provided data ($R^2 = 0.0039$).

282

283 **[Table 1]**

284

285 *Visual assessment*

286 During trot away from the observer, the percentage viewing time was significantly different
287 between the two cohorts for regions 5, 6, 7, 8, 9, 13 and 14 ($P \leq 0.041$), but not for the
288 remaining regions ($P \geq 0.114$). During trot towards the observer, the percentage viewing time
289 was significantly different between the two cohorts for regions 6, 7, 12, 13 and 14 ($P \leq$
290 0.047), but not for the remaining regions ($P \geq 0.075$). Differences in distribution of gaze data
291 are illustrated in Figure 5. There was considerable variation across students in the distribution
292 of time across body regions for both trotting directions.

293

294 *Questionnaire responses*

295 The inexperienced cohort rated the difficulty of performing a lameness examination (Table 2)
296 with a mean (SD) score of 4.10 (0.48) out of 5, ranging from 3 to 5. The experienced cohort

297 rated the difficulty with a score of 3.35 (0.47) out of 5, ranging from 3 to 4. The median [IQR]
298 number of lameness cases seen during their time at university by the experienced cohort was
299 24 [30] cases, ranging from 6 to 100. Assessment away from / towards the observer was
300 judged most difficult by the majority of inexperienced students (12 / 20 students, 60%), while
301 assessment away from / towards the observer (5 / 12 students, 42%) and in lateral view (6 / 12
302 students, 50%) were ranked as most difficult by a comparable number of experienced
303 students.

304

305 **[Table 2]**

306

307 Compared to the calculated performance (Figure 6), students underestimated their skill by -11
308 (15) % (inexperienced cohort) and -13 (12) % (experienced cohort). The difference between
309 self-rating and objective assessment of performance was not significantly different between
310 cohorts.

311

312 **[Figure 6]**

313

314 Both, the inexperienced and experienced cohort, most frequently named the use of head
315 movement for the detection of forelimb lameness on the straight line and circle. Both groups
316 also most frequently named pelvis movement on the straight and stride length on the circle for
317 the detection of hind limb lameness (Tables 3 and 4). The inexperienced cohort named a total
318 of 19 and 18 features for forelimb lameness assessment on straight and circle, respectively,
319 compared to 9 and 6 features named by the experienced cohort. The inexperienced cohort

320 named a total of 16 and 21 features for hind limb lameness assessment on straight and circle,
321 respectively, compared to 5 and 7 features named by the experienced cohort.

322

323 **[Table 3 and 4]**

324

325 **Discussion**

326 *Student lameness assessment skills*

327 This study showed that both inexperienced and experienced student cohorts, while in general
328 reliably classifying horses as lame, were weak in correctly determining the affected limb, and
329 only performed at the level of chance when classifying a horse as sound. The percentage of
330 horses classified as lame in the correct limb significantly differed between the two cohorts,
331 highlighting what we hope is the effect of learning throughout the veterinary degree.
332 However, the variation in performance across experienced students was not explained by the
333 amount of lame horses that they had previously observed. This finding substantiates the view
334 that just looking at many horses does not necessarily improve skill level. Rather, the amount
335 of time engaged in deliberate practice, the act of practising and refining a skill actively,
336 accounts for most of the expertise level accomplished (Ericsson and others 1993). Even for
337 the experienced cohort, performance was still not at a level which we would deem acceptable
338 for reliable and repeatable diagnoses: there was large variation in performance across students
339 and a median of less than 70 % correctly classified horses, with the exception of forelimb
340 lameness on the circle. Consequently, students close to graduation would on average still
341 incorrectly assess one out of four horses on the straight line, with individual participants
342 assessing only 30 % of horses correctly. For the inexperienced cohort, individual performance
343 was as low as 0% correct. If these results generalise to other veterinary students, it is clearly

344 important to ensure more targeted teaching of lameness assessment skills, to improve student
345 performance, especially of those choosing an equine career path. Future training should
346 specifically target detection of mild lameness, a task difficult even for expert assessors
347 (Keegan and others 2010; Keegan and others 2012). Also, since most mistakes resulted from
348 classifying the incorrect limb as lame, an emphasis should be placed on how to recognise the
349 correct limb.

350 We found differences in performance between assessment on straight line and circle as well
351 as between fore- and hind limb lameness. Only on the circle did experienced students show an
352 outstanding performance for the detection of forelimb lameness with a median of 100 %. On
353 the circle, the mean lameness score derived from clinical records was higher (3.67) compared
354 to the straight line (2.38). This is likely to have made lameness easier to see: with increasing
355 lameness, the head nod converges on almost a single vertical head excursion per stride
356 (Clayton 1987; Peloso and others 1993; Christovão and others 2007), resulting in an
357 unambiguous nod-down pattern. This pattern is easier to see than differences in the two
358 excursions between the two steps of a stride during milder lameness, since the eye has to
359 follow movement at only have the step frequency and there is only a single minimum and
360 maximum during the whole stride. In line with findings for more experienced clinicians
361 (Keegan and others 2010), agreement amongst students was lowest for the presence or
362 absence of hind limb lameness on both, straight line and circle. In this study, the variation in
363 assigned scores across students compared well with previous findings, where final year
364 students graded no more than three grades away from a horse's median grade, while more
365 experienced assessors graded no more than two grades and experts graded no more than one
366 grade away (Arkell and others 2006).

367 *Reliability of self-evaluation and sources of error*

368 While students did not perform at a competent level, their self-evaluation reflected their
369 ‘objective’ performance, showing a realistic view of their own competence, even a slight
370 underestimate. This finding is a positive sign, for their safety as practitioners, as they are
371 recognising their own limitations while it also marks a positive sign for learning. The self-
372 evaluation of experienced medical practitioners often does not bear any correlation with
373 actual capabilities (Tracey and others 1997), a finding also observed when reviewing self-
374 assessments of inexperienced and experienced medical personal (Gordon 1991). However,
375 increasing self-awareness may lead to better self-evaluation (Mabe and West 1982) and
376 identification of the need for learning. Inexperienced students, in particular, repeatedly
377 mentioned struggling with the task due to the lack of learning opportunities and uncertainty
378 regarding lameness features.

379 Errors in classifying the lame limb were rooted in what we consider to be both conceptual and
380 perceptual mistakes. Selection of the incorrect contralateral limb (left vs. right limb) is likely
381 to have been related to a conceptual error, where the student did not correctly remember the
382 rules relating a movement feature to the lame limb. In contrast, the selection of the incorrect
383 ipsilateral limb (forelimb vs. hind limb) may be related to a perceptual error due to the
384 interaction of forelimb and hind limb lameness: especially for hind limb lameness,
385 compensatory forelimb movements are a regular and notable occurrence (Uhlir and others
386 1997; Kelmer and others 2005). For forelimb lameness, compensatory hind limb movements
387 develop only with more pronounced lameness and to a lesser visible extent (Uhlir and others
388 1997; Kelmer and others 2005). This linkage between the fore and hind limb movements may
389 lead the student to mistaking compensatory movement for primary lameness, distracting them
390 from further lameness evaluation. In future, study designs should account for this source of
391 error by explicitly requesting observers to evaluate all limbs and if more than one appears
392 lame, to indicate whether the lameness is assumed to be primary or “compensatory”.

393 Students performed at chance level for the classification of sound horses as non-lame. This
394 finding highlights the need to not only train students in recognising lameness, but also provide
395 training in recognising soundness and familiarising students with the spectrum of movement
396 in sound horses. Knowing what is ‘normal’ and what is not is a key element for the
397 development of discrimination skills, which in turn is an important component of expertise
398 development (Kellman and Garrigan 2009). Research into the teaching of radiographic
399 interpretation has shown that allowing students to compare abnormal and normal radiographs
400 during training resulted in better performance in a subsequent detection test compared to
401 students who were trained with abnormal radiographs only (Kok and others 2013). Similarly,
402 facilitating ‘learning by contrast’ or ‘comparison learning’ (Kurtz and Boukrina 2005;
403 Gentner 2010), for students learning how to analyse electrocardiograms, had a positive effect
404 on subsequent student performance of the task (Hatala and others 2003; Ark and others 2007).
405 The need for greater standardisation of gait analysis training and a greater emphasis on what
406 is normal to improve inter-rater reliability has equally been highlighted for human gait
407 analysis (Eastlack and others 1991). Hence, in future, it will be advisable to incorporate a
408 greater number of sound horses into student training.

409 *Visual assessment strategies*

410 The mapped eye tracking data showed that the cohort of inexperienced students dedicated
411 more time to the assessment of limb movement compared to the experienced cohort, both with
412 the horse moving away from and towards the observer. The experienced students dedicated
413 most time to the assessment of the pelvis area (trot away from the observer) and the head area
414 (trot towards the observer). These findings matched the lameness features named by the two
415 student cohorts: the inexperienced cohort named 16 to 21 features for assessment of fore- and
416 hind limb lameness on straight and circle, many of them relating to movement of the limbs. In
417 contrast, the experienced cohort named only 5 to 9 features across conditions, a large
418 proportion referring to movement of head and pelvis. These findings indicate that as part of

419 the development of diagnostic expertise, students may discard redundant lameness features
420 and rely on the few features which they assume to be most reliable. However, as a possible
421 confounding factor, experienced students may have been taught by the same veterinarian as
422 part of their clinical lameness rotations, which could have caused the shift in feature selection.
423 Variation in the distribution of gaze data and described lameness features between students of
424 the same cohort as well as between cohorts highlighted inconsistency in assessment approach
425 and evaluation protocol, which should be addressed in future training especially for students
426 early in their careers to ensure that they acquire a more consistent skill set.

427

428 **Conclusions**

429 In this study we investigated veterinary student competency in lameness evaluation prior to
430 and during clinical rotations. We found that the performance level of both cohorts was
431 inadequate in most scenarios, where mistakes largely arose from classifying an incorrect limb
432 as lame. We further found substantial differences in the assessment strategy, where the
433 experienced student cohort primarily looked at upper body movement (head and sacrum) and
434 named only few lameness pointers, while the inexperienced cohort focused to a large extent
435 on limb movement and named a large variety of lameness pointers. We conclude that students
436 require a higher level of perceptual training before commencing clinical rotations in order to
437 clarify relevant lameness pointers and acquire the necessary discriminatory skill allowing
438 them to differentiate sound from lame horses and correctly pinpoint the affected limb(s). We
439 have by now released a free online training tool which aims to develop these skills, which can
440 be found at www.lamenesstrainer.com. We hope that in future, systematic training based on
441 the findings presented here, which takes into consideration common misconceptions and -
442 possibly incorrect- 'intuitive' approaches, may help students to develop more reliable
443 lameness assessment skills.

444

445 **Acknowledgements**

446 We would like to thank Jon Ward of Acuity ETS for supporting access to the eye tracker for
447 the duration of the study, Prof. Roger K Smith for the video recordings of clinical cases, Thilo
448 Pfau and Renate Weller for help with participant recruitment and all of our participants for
449 their time. S. D. Starke's PhD was funded by the Mellon Trust *via* the Royal Veterinary
450 College, UK.

451

452 **References**

- 453 ARK, T. K., BROOKS, L. R., EVA, K. W. (2007). The benefits of flexibility: the pedagogical
454 value of instructions to adopt multifaceted diagnostic reasoning strategies. *Medical*
455 *Education* **41**, 281-287.
- 456 ARKELL, M., ARCHER, R. M., GUITIAN, F. J., MAY, S. A. (2006). Evidence of bias
457 affecting the interpretation of the results of local anaesthetic nerve blocks when
458 assessing lameness in horses. *Vet Rec.* **159**, 346-348.
- 459 BALSLEV, T., JARODZKA, H., HOLMQVIST, K., DE GRAVE, W., MUIJTJENS, A. M.
460 M., EIKA, B., VAN MERRIËNBOER, J., SCHERPBIER, A. J. J. A. (2012). Visual
461 expertise in paediatric neurology. *European Journal of Paediatric Neurology* **16**, 161-
462 166.
- 463 BRENNAN, R. L., PREDIGER, D. J. (1981). Coefficient Kappa - Some Uses, Misuses, and
464 Alternatives. *Educational and Psychological Measurement* **41**, 687-699.
- 465 BROWN, C. R., HILLMAN, S. J., RICHARDSON, A. M., HERMAN, J. L., ROBB, J. E.
466 (2008). Reliability and validity of the Visual Gait Assessment Scale for children with
467 hemiplegic cerebral palsy when used by experienced and inexperienced observers.
468 *Gait & Posture* **27**, 648-652.
- 469 BRUNNEKREEF, J. J., VAN UDEN, C. J., VAN MOORSEL, S., KOOLOOS, J. G. (2005).
470 Reliability of videotaped observational gait analysis in patients with orthopedic
471 impairments. *BMC Musculoskelet Disord* **6**, 17.
- 472 CHRISTOVÃO, F. G., BARROS, R. M. L., MARTINS, C. B., SILVA, M. A. G., FREITAS,
473 E. V. V., LACERDA NETO, J. C., QUEIROZ-NETO, A. (2007). Three-dimensional
474 kinematic analysis of head and limb movements of lame and non-lame colts. *Equine*
475 *and Comparative Exercise Physiology* **4**, 31-36.
- 476 CLAYTON, H. M. (1987). Cinematographic analysis of the gait of lame horses III: Fracture
477 of the third carpal bone. *Journal of Equine Veterinary Science* **7**, 130-135.
- 478 DOW, S. M., LEENDERTZ, J. A., SILVER, I. A., GOODSHIP, A. E. (1991). Identification
479 of subclinical tendon injury from ground reaction force analysis. *Equine Veterinary*
480 *Journal* **23**, 266-272.
- 481 EASTLACK, M. E., ARVIDSON, J., SNYDER-MACKLER, L., DANOFF, J. V.,
482 MCGARVEY, C. L. (1991). Interrater reliability of videotaped observational gait-
483 analysis assessments. *PHYS THER* **71**, 465-72.
- 484 ERICSSON, K. A., KRAMPE, R. T., TESCH-RÖMER, C. (1993). The role of deliberate
485 practice in the acquisition of expert performance. *Psychological Review* **100**, 363-406.
- 486 FINDLAY, J. M., GILCHRIST, I. D. (2003). *Active Vision - The psychology of looking and*
487 *seeing*. Oxford, Oxford University Press.

- 488 GENTNER, D. (2010). Bootstrapping the mind: analogical processes and symbol systems.
489 *Cogn Sci* **34**, 752-75.
- 490 GORDON, M. J. (1991). A review of the validity and accuracy of self-assessments in health
491 professions training. *Academic Medicine* **66**, 762-9.
- 492 HATALA, R., BROOKS, L., NORMAN, G. (2003). Practice Makes Perfect: The Critical
493 Role of Mixed Practice in the Acquisition of ECG Interpretation Skills. *Advances in*
494 *Health Sciences Education* **8**, 17-26.
- 495 ISHIHARA, A., BERTONE, A. L., RAJALA-SCHULTZ, P. J. (2005). Association between
496 subjective lameness grade and kinetic gait parameters in horses with experimentally
497 induced forelimb lameness. *American Journal of Veterinary Research* **66**, 1805-1815.
- 498 KAWAMURA, C. M., DE MORAIS FILHO, M. C., BARRETO, M. M., DE PAULA ASA,
499 S. K., JULIANO, Y., NOVO, N. F. (2007). Comparison between visual and three-
500 dimensional gait analysis in patients with spastic diplegic cerebral palsy. *Gait &*
501 *Posture* **25**, 18-24.
- 502 KEEGAN, K. G., DENT, E. V., WILSON, D. A., JANICEK, J., KRAMER, J.,
503 LACARRUBBA, A., WALSH, D. M., CASSELLS, M. W., ESTHER, T. M.,
504 SCHILTZ, P., FREES, K. E., WILHITE, C. L., CLARK, J. M., POLLITT, C. C.,
505 SHAW, R., NORRIS, T. (2010). Repeatability of subjective evaluation of lameness in
506 horses. *Equine Vet J* **42**, 92-7.
- 507 KEEGAN, K. G., MACALLISTER, C. G., WILSON, D. A., GEDON, C. A., KRAMER, J.,
508 YONEZAWA, Y., MAKI, H., PAI, P. F. (2012). Comparison of an inertial sensor
509 system with a stationary force plate for evaluation of horses with bilateral forelimb
510 lameness. *American Journal of Veterinary Research* **73**, 368-74.
- 511 KEEGAN, K. G., WILSON, D. A., WILSON, D. J., SMITH, B., GAUGHAN, E. M.,
512 PLEASANT, R. S., LILLICH, J. D., KRAMER, J., HOWARD, R. D., BACON-
513 MILLER, C., DAVIS, E. G., MAY, K. A., CHERAMIE, H. S., VALENTINO, W. L.,
514 VAN HARREVELD, P. D. (1998). Evaluation of mild lameness in horses trotting on
515 a treadmill by clinicians and interns or residents and correlation of their assessments
516 with kinematic gait analysis. *American Journal of Veterinary Research* **59**, 1370-
517 1377.
- 518 KELLMAN, P. J., GARRIGAN, P. (2009). Perceptual learning and human expertise. *Physics*
519 *of Life Reviews* **6**, 53-84.
- 520 KELMER, G., KEEGAN, K. G., KRAMER, J., WILSON, D. A., PAI, F. P., SINGH, P.
521 (2005). Computer-assisted kinematic evaluation of induced compensatory movements
522 resembling lameness in horses trotting on a treadmill. *American Journal of Veterinary*
523 *Research* **66**, 646-655.
- 524 KESTIN, S. C., KNOWLES, T. G., TINCH, A. E., GREGORY, N. G. (1992). Prevalence of
525 Leg Weakness in Broiler-Chickens and Its Relationship with Genotype. *Veterinary*
526 *Record* **131**, 190-194.
- 527 KOK, E. M., DE BRUIN, A. B. H., ROBBEN, S. G. F., VAN MERRIËNBOER, J. J. G.
528 (2013). Learning radiological appearances of diseases: Does comparison help?
529 *Learning and Instruction* **23**, 90-97.
- 530 KRUPINSKI, E. A. (1996). Visual scanning patterns of radiologists searching mammograms.
531 *Academic Radiology* **3**, 137-144.
- 532 KRUPINSKI, E. A. (2005). Visual Search of Mammographic Images: Influence of Lesion
533 Subtlety1. *Academic Radiology* **12**, 965-969.
- 534 KRUPINSKI, E. A. (2010). Current perspectives in medical image perception. *Attention,*
535 *Perception, & Psychophysics* **72**, 1205-1217.
- 536 KUNDEL, H. L., NODINE, C. F., CARMODY, D. (1978). Visual Scanning, Pattern
537 Recognition and Decision-making in Pulmonary Nodule Detection. *Investigative*
538 *Radiology* **13**, 175-181.

- 539 KURTZ, K. J., BOUKRINA, O. (2005). Learning Relational Categories by Comparison of
540 Paired Examples. *Proceedings of the Twenty-Sixth Annual Conference of the*
541 *Cognitive Science Society*, 756-761.
- 542 MAATHUIS, K. G. B., VAN DER SCHANS, C. P., VAN IPEREN, A., RIETMAN, H. S.,
543 GEERTZEN, J. H. B. (2005). Gait in children with cerebral palsy - Observer
544 reliability of Physician Rating Scale and Edinburgh Visual Gait Analysis Interval
545 Testing Scale. *Journal of Pediatric Orthopaedics* **25**, 268-272.
- 546 MABE, P. A., WEST, S. G. (1982). Validity of self-evaluation of ability: A review and meta-
547 analysis. *Journal of Applied Psychology* **67**, 280-296.
- 548 MAIN, D. C. J., CLEGG, J., SPATZ, A., GREEN, L. E. (2000). Repeatability of a lameness
549 scoring system for finishing pigs. *Vet Rec.* **147**, 574-576.
- 550 MC CRACKEN, M. J., KRAMER, J., KEEGAN, K. G., LOPES, M., WILSON, D. A.,
551 REED, S. K., LACARRUBBA, A., RASCH, M. (2012). Comparison of an inertial
552 sensor system of lameness quantification with subjective lameness evaluation. *Equine*
553 *Veterinary Journal* **44**, 652-656.
- 554 ONG, A. M. L., HILLMAN, S. J., ROBB, J. E. (2008). Reliability and validity of the
555 Edinburgh Visual Gait Score for cerebral palsy when used by inexperienced observers.
556 *Gait & Posture* **28**, 323-326.
- 557 ORITO, K., KUROZUMI, S., ISHII, I., TAKAHASHI, T., HIRAGA, A., MATSUDA, H.
558 (2007). Identification of Equine Subclinical Lameness Induced by Pressure to the Sole
559 of Fore- or Hindlimb. *Journal of Equine Veterinary Science* **27**, 429-434.
- 560 PELOSO, J. G., STICK, J. A., SOUTAS-LITTLE, R. W., CARON, J. C., DECAMP, C. E.,
561 LEACH, D. H. (1993). Computer-assisted three-dimensional gait analysis of
562 amphotericin-induced carpal lameness in horses. *American Journal of Veterinary*
563 *Research* **54**, 1535-43.
- 564 RANDOLPH, J. J. (2005). Free-marginal multirater kappa: An alternative to Fleiss' fixed-
565 marginal multirater kappa. *Joensuu Learning and Instruction Symposium 2005*.
566 Joensuu, Finland.
- 567 READ, H. S., HAZLEWOOD, M. E., HILLMAN, S. J., PRESCOTT, R. J., ROBB, J. E.
568 (2003). Edinburgh visual gait score for use in cerebral palsy. *Journal of Pediatric*
569 *Orthopaedics* **23**, 296-301.
- 570 ROSS, M. W., DYSON, S. J. (2011). *Diagnosis and Management of Lameness in Horses*. ,
571 St. Louis, Missouri, Elsevier Saunders.
- 572 STASHAK, T. S. (2002). *Adams' Lameness in Horses*, Wiley-Blackwell.
- 573 STASHAK, T. S., HILL, C. (1996). *Practical Guide to Lameness in Horses*, Wiley-
574 Blackwell.
- 575 THOMAS, E. L., LANSDOWN, E. L. (1963). Visual Search Patterns of Radiologists in
576 Training. *Radiology* **81**, 288-292.
- 577 TIEN, T., PUCHER, P. H., SODERGREN, M. H., SRISKANDARAJAH, K., YANG, G.-Z.,
578 DARZI, A. Eye tracking for skills assessment and training: a systematic review.
579 *Journal of Surgical Research* **191**, 169-178.
- 580 TRACEY, J. M., ARROLL, B., RICHMOND, D. E., BARHAM, P. M. (1997). The validity
581 of general practitioners' self assessment of knowledge: cross sectional study. *BMJ* **315**,
582 1426-8.
- 583 UHLIR, C., LICKA, T., KUBBER, P., PEHAM, C., SCHEIDL, M., GIRTTLER, D. (1997).
584 Compensatory movements of horses with a stance phase lameness. *Equine Vet J Suppl*
585 **23** **29**, 102-5.
- 586 WARRENS, M. (2010). Inequalities between multi-rater kappas. *Advances in Data Analysis*
587 *and Classification* **4**, 271-286.
- 588 WAXMAN, A. S., ROBINSON, D. A., EVANS, R. B., HULSE, D. A., INNES, J. F.,
589 CONZEMIUS, M. G. (2008). Relationship Between Objective and Subjective

590 Assessment of Limb Function in Normal Dogs with an Experimentally Induced
591 Lameness. *Veterinary Surgery* **37**, 241-246.
592 WILLIAMS, G., MORRIS, M. E., SCHACHE, A., MCCRORY, P. (2009). Observational gait
593 analysis in traumatic brain injury: Accuracy of clinical judgment. *Gait & Posture* **29**,
594 454-459.
595 WYN-JONES, G. (1988). Chapter 1. The Diagnosis of the Causes of Lameness. *In*: WYN-
596 JONES, G. (ed.) *Equine Lameness*. Oxford: Blackwell Scientific Publications.
597
598

599

600 **TABLES**

601

602 **Table 1. Within-observer agreement for the two student cohorts.**

603 Free marginal kappa for the binary classification of horses as lame (yes/no), forelimb lame
 604 (yes/no) and hind limb lame (yes/no) for trotting on a straight line away from/towards the
 605 observer and in a circle.

		Inexperienced students	Experienced students
Lame?	Away / towards	0.54	0.49
	Circle	0.70	0.89
Forelimb lame?	Away / towards	0.30	0.64
	Circle	0.51	0.67
Hind limb lame?	Away / towards	0.15	0.20
	Circle	0.25	0.28

606

607

608 **Table 2. Difficulty ranking for assessing different views.**

609 Count of students ranking the views ‘straight line’, ‘circle’ and ‘lateral’ as 1 (easiest), 2
 610 (medium) and 3 (most difficult). One experienced student perceived assessment on straight and
 611 circle to be of identical difficulty.

Difficulty ranking	Inexperienced students			Experienced students		
	<i>Straight</i>	<i>Circle</i>	<i>Lateral</i>	<i>Straight</i>	<i>Circle</i>	<i>Lateral</i>
Easiest (1)	5	8	7	7	4	1
Medium (2)	3	10	7	0	7	6
Most difficult (3)	12	2	6	5	1	6

612

613

614

615 **Table 3. Named visual forelimb lameness pointers.**

616 Pointers described for the determination of the presence of forelimb lameness during trot on the
 617 straight line and in a circle by respondents to the questionnaire for the cohort of inexperienced
 618 students (N = 19) and experienced students (N = 12). **: most commonly named feature, *:
 619 second most commonly named feature.

Lameness pointer	Inexperienced students		Experienced students	
	Straight	Circle	Straight	Circle
Head movement	18**	16**	12**	12**
Foot placement	6*	3	-	-
Circumduction	4	3	1	-
Stride / Step length	3	8*	4*	8*
Stance time	2	1	1	2
Toe dragging	2	1	-	-
Difference in load bearing	1	2	1	1
Lateral bending of head and neck	1	-	-	-
Shoulders	1	1	1	-
Evenness of striding / uneven gait	1	1	-	-
Dropping elbow	1	-	-	-
Hoof flight arc	1	1	-	-
Limb stiffness / general stiffness	1	1	-	-
Longer time with 1 foot in air	1	-	1	-
Hypermetria (abnormality in stepping)	1	1	-	-
Muscle atrophy	1	-	-	-
Tracking up	1	-	-	-
Fetlock extension	1	1	1	1
Loading of carpus	1	-	-	-
Looking out with affected limb on inside	-	2	-	-
Limb swing	-	1	-	-
magnitude of bending	-	1	-	-
Rhythm	-	1	-	-
Hopping sort of gait	-	1	-	-
Axillae movement asymmetry	-	-	2	-
Uncomfortable, e.g. not going forward, stopping	-	-	-	1

620

621

622 **Table 4. Named visual hind limb lameness pointers.**

623 Pointers described for the determination of the presence of hind limb lameness as detailed in

624 Table 3, **: most commonly named feature, *: second most commonly named feature.

625

Lameness pointer	Inexperienced		Experienced	
	Straight	Circle	Straight	Circle
Hind quarters movement	15**	6*	12**	9*
Stride length	5*	9**	2*	10**
Foot placement	4	3		
Toe dragging	3	1		
Circumduction	3	2		
Stance time	2	2	1	1
Level of hocks	2			
Muscle conformation / atrophy	2	1		
Head nod	2	3		
Fetlock extension	2		2*	3
Limb stiffness	1	1		
How the legs stride through	1	1		
Less head nodding	1			
Rump moving up and down more, bouncing on one leg	1	1		
Hypermetria (abnormality in stepping)	1			
Limb (range of) movement	1			1
Tracking up		3		
Foot flight arc		2	1	
Stiffness through the back		1		
Don't know		1		
Reluctance to bend and limb position		1		
magnitude of bending		1		
Leg angles		1		1
Rhythm		1		
Freedom of movement through hip		1		
Not wanting to keep on the circle		1		
Speed of protraction				1

626

627

628

629 **FIGURE LEGENDS**

630

631 **Figure 1.** Different views in which horses were presented in the video clips: trot on a straight
632 line away from the observer (a) and towards the observer (b), trot in a circle (c) and trot in
633 lateral view (d).

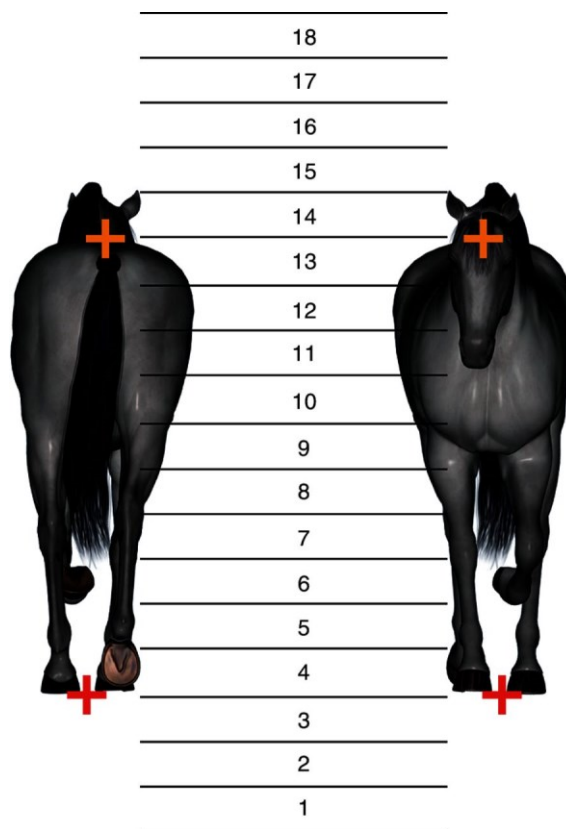


634

635

636

637 **Figure 2.** Mapping of eye tracking data for all videos presenting the horse during trot away
638 from / towards the observer. Illustration of grid overlay to map the point of gaze to discrete
639 regions of interest, respectively bins (1-18). The grid was evenly spaced and calculated based
640 on the location of strategic features for each frame: based on the location of the os sacrum
641 (orange cross) and feet (red cross) for trot away from the observer (left), and the location of
642 the head (orange cross) and feet (red cross) for trot towards the observer (right). The two
643 points were used to create a grid of regular spacing, scaled to the height of the horse in each
644 respective frame.

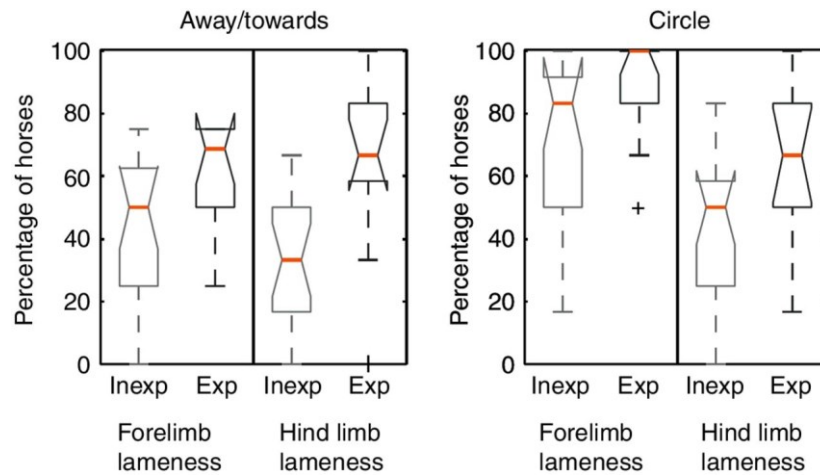


645

646

647

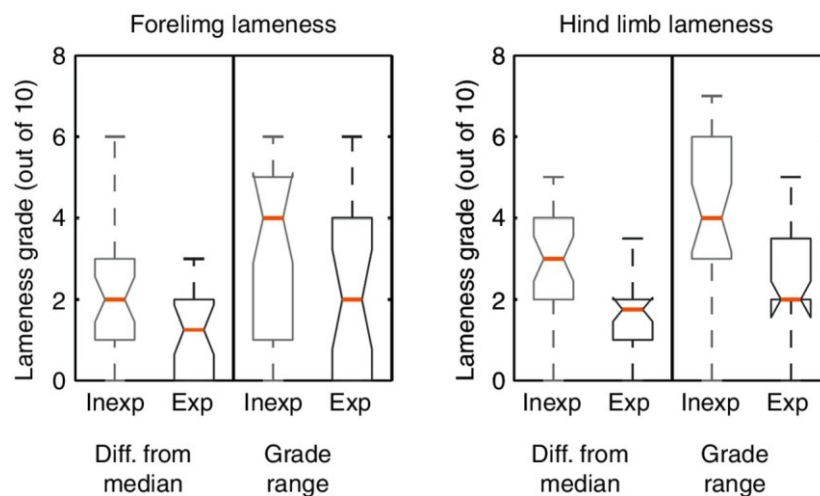
648 **Figure 3.** Boxplots of the percentage of lame horses classed as lame in the correct limb by a
 649 cohort of 20 inexperienced students ('Inexp') as well as 12 experienced students ('Exp')
 650 during assessment of horses at trot on the straight line away from/towards the observer (left)
 651 and at trot in a circle (right). Orange line: median, +: outlier.



652

653

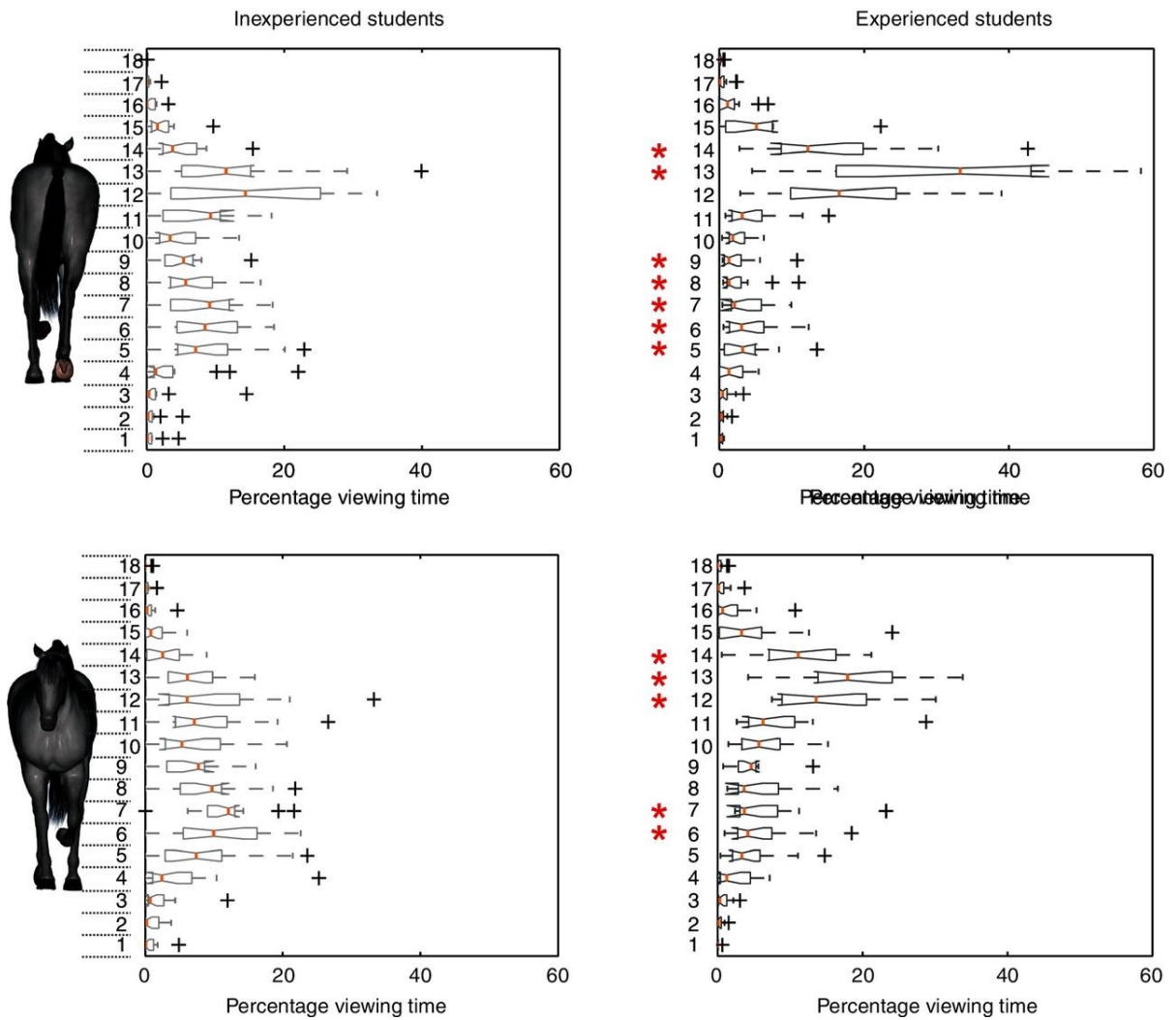
654 **Figure 4.** Boxplots of the variation in lameness grades assigned to horses by students of the
 655 inexperienced ('Inexp') and experienced ('Exp') cohort. For each cohort, the difference from
 656 the cohort's median grade as well as the total score range were calculated for those horses
 657 which each individual student considered lame with the most affected limb being a forelimb
 658 (left) or hind limb (right). Orange line: median.



659

660

661 **Figure 5.** Boxplots for the percentage viewing time spent on each of the 18 discrete image
 662 regions based on the average value calculated for each student of the inexperienced (left) and
 663 experienced (right) cohort during trot away from the observer (top) and towards the observer
 664 (bottom). Orange line: median, +: outlier.

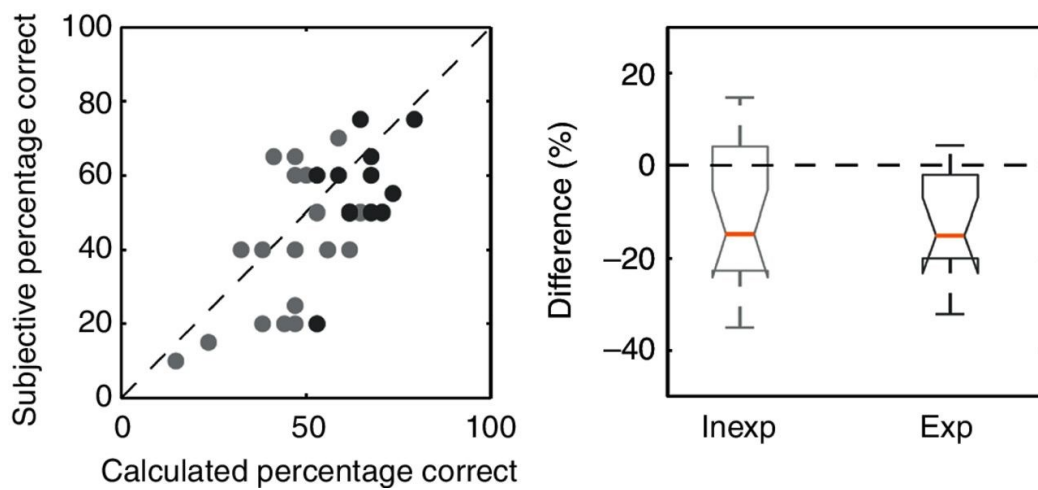


665

666

667

668 **Figure 6.** Left: comparison of subjective (self-rating by the student) and objective (calculated
669 percentage correct when compared to clinical ground truth) performance evaluation. The
670 diagonal line indicates an exact match between the the student's self evaluation and actual
671 performance. Light grey: inexperienced students, dark grey: experienced students. Right:
672 Difference between subjective and objective performance estimate for the inexperienced
673 cohort ('Inexp) and the experienced cohort ('Exp). Most differences are below zero,
674 indicating that the self-assessment was rather conservative.



675