

Online Radiology Reporting with Peer Review as a Learning and Feedback Tool in Radiology; Implementation, Validity, and Student Impressions

Fintan J. McEvoy¹ · Nicholas W. Shen^{1,2} · Dorte H. Nielsen¹ · Lene E. Buelund¹ · Peter Holm¹

Published online: 3 October 2016
© Society for Imaging Informatics in Medicine 2016

Abstract Communicating radiological reports to peers has pedagogical value. Students may be uneasy with the process due to a lack of communication and peer review skills or to their failure to see value in the process. We describe a communication exercise with peer review in an undergraduate veterinary radiology course. The computer code used to manage the course and deliver images online is reported, and we provide links to the executable files. We tested to see if undergraduate peer review of radiological reports has validity and describe student impressions of the learning process. Peer review scores for student-generated radiological reports were compared to scores obtained in the summative multiple choice (MCQ) examination for the course. Student satisfaction was measured using a bespoke questionnaire. There was a weak positive correlation (Pearson correlation coefficient = 0.32, $p < 0.01$) between peer review scores students received and the student scores obtained in the MCQ examination. The difference in peer review scores received by students grouped according to their level of course performance (high vs. low) was statistically significant ($p < 0.05$). No correlation was found between peer review scores awarded by the students and the scores they obtained in the MCQ examination (Pearson correlation coefficient = 0.17, $p = 0.14$). In conclusion, we have created a realistic radiology imaging exercise

with readily available software. The peer review scores are valid in that to a limited degree they reflect student future performance in an examination. Students valued the process of learning to communicate radiological findings but do not fully appreciate the value of peer review.

Keywords Radiology teaching files · Education–Medical · Programming languages

Introduction

Learning to communicate in one's specialist domain fosters a sense of identity. It is central to social theories of learning. The process of writing and speaking like a radiologist or medical doctor alters the student's self-perception [1]. Peer review in the learning process fosters a sense of responsibility and ownership for learning among peers, enhances skills in self-assessment [2], promotes self-learning, improving self-confidence [3], and provides experience for the workplace [4]. Reported obstacles to communication and peer review in radiology relate to implementation [1, 5], student perceptions of validity [4], and lack confidence in performing reviews [6]. Outside the sphere of teaching and learning, peer review is used as a quality control tool in radiology departments [7, 8]. As a learning tool and for its own sake, student exposure to peer review is important.

We introduced a communication and peer review exercise to improve student satisfaction with and involvement in a 5-week veterinary imaging course. We hoped to achieve this without increasing teachers' workload and by using available open source or free software to its maximum advantage. Satisfaction with image reading exercises in the existing course was suboptimal according to previous formal and informal post course student feedback. The image reporting

✉ Fintan J. McEvoy
fme@sund.ku.dk

¹ Department of Veterinary Clinical and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Dyrhøjevej 16, 1870 Frederiksberg C, Denmark

² Present address: College of Veterinary Medicine, Lincoln Memorial University, 6965 Cumberland Gap Parkway, Harrogate, TN 37752, USA

exercise was found to be a source of frustration for the students. Problems identified included students remaining passive in peer groups, unwillingness to commit or form opinions about the images, and a desire to constantly seek input from teaching staff in preference to independent research and thinking.

The imaging exercise described here was introduced in response to these problems and tailored to the course's needs. Previous experience with peer review in radiology teaching with more senior students had been positive [9]. For the current implementation, student performance and satisfaction was monitored.

Our experiences with the implementation and assessment of peer review in a veterinary radiology course are reported here. In particular, we report the logistics of providing web access to radiology images for teaching and we examined the skill of students in performing peer review, the possibility of differences in peer review performance according to the overall level of performance in the course, and the students' level of satisfaction with the material and peer review. We hypothesized that online delivery would be realistic and convenient, that student performance in peer review would correlate with performance in the summative course examination, and that students would see value in the peer review process.

Methods

Access to radiological images (radiographic, magnetic resonance imaging, and computed tomography) for student reporting with peer review was introduced and continues in use as a 6-day module within a 5-week veterinary radiology course. It is placed at the end of the course, which typically comprises 20 to 22 students, all of whom will have completed courses in basic anatomy and pathology. The course runs four times per year, and data from 82 students participating in four courses are reported here. It covers basic undergraduate requirements in veterinary imaging, radiation safety, radiographic technique, interpretation, and radiographic anatomy. In preparation for the case review module, students are instructed on how to write formal structured radiographic reports. Two to 4 days prior to the start of the module, one of the course leaders uses a 40-min session to tell the students the purposes of the module, how to access case material, how to enter reports, and how to perform peer reviews. For the last of these, students are told how to access a peer review form, informed that it comprises a series of check boxes, and told when peer review is to be performed. The peer review form asks the extent to which essential components of a radiographic report were present (the peer review questions are provided below). In addition, students were told that the performance of peer review is voluntary, but the honorable course of action is to perform careful peer review. No formal information was

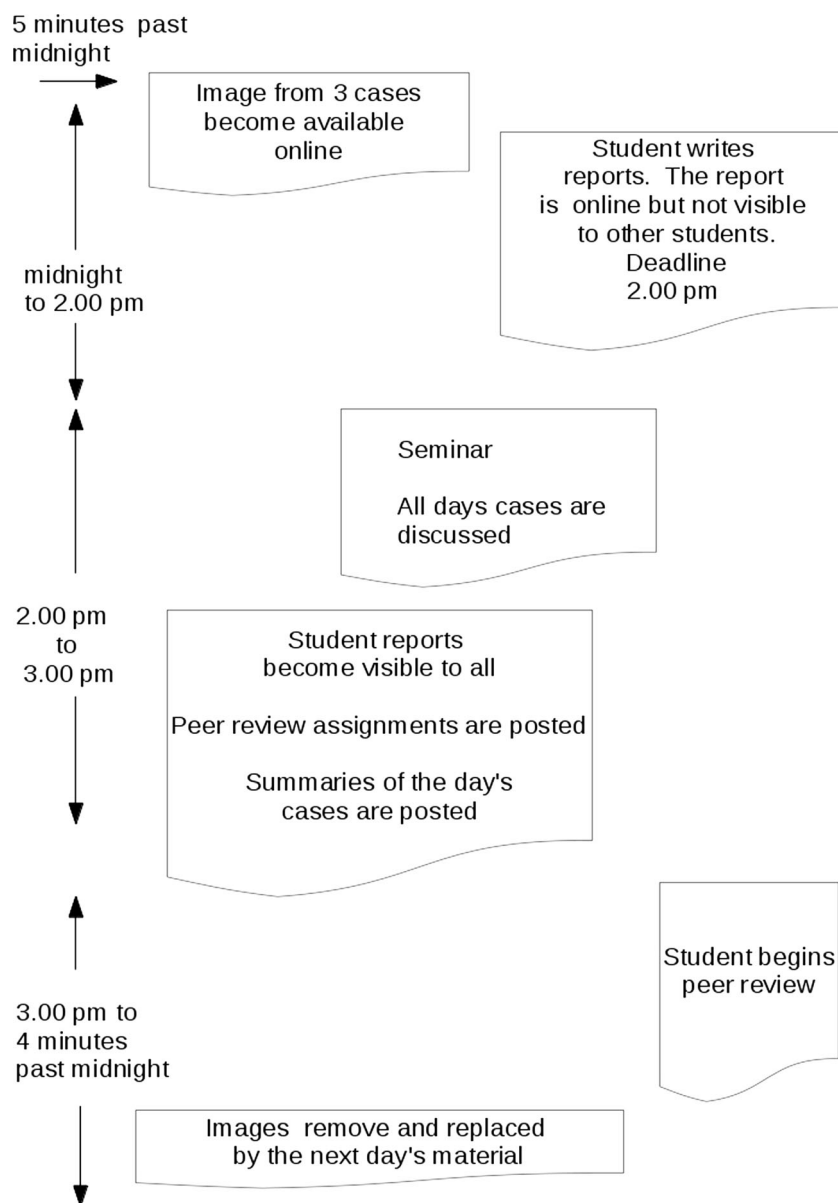
provided on the pedagogic merits and rationale of peer review. The course concludes with a summative multiple choice examination (MCQ), which contains approximately 40 questions that cover the entire course content. Performance or activity in peer review is not factored into the final score for the course. Students are informed that participation in the image reporting and peer review process is formative and not summative.

Timeline, Independent Study, and Group Sessions The 6-day case reading session with peer review runs to a fixed time line. To mimic a real life situation, where reading and interpretation of radiographs is needed within a limited time span, the cases for each day automatically become available at 5 min past midnight and are available for 24 h only. Students are told to write independent reports and submit them online. They are instructed not to work in groups. They are free to ask general questions as they arise for the cases, the guideline being that if the information needed is available in a textbook, one is free to ask a colleague, perform an internet search, or use textbooks to find an answer. Students are instructed not to ask or to say the particular diagnosis or a differential diagnosis list for a particular case. A dedicated chat line (Kiwi IRC. <https://kiwiiec.com>) is available on a web site for students to communicate, while reporting, if they wish. The case reports are to be completed by 2:00 pm each day. At that time, all students taking the course (typically 20 in number) meet with one of the teachers for a 1-h seminar where each case is displayed, reported orally by one of the students and discussed. At the end of this session at 3:00 pm, the teaching points for the day's cases and the peer review assignments are automatically updated on the course web page. In addition, viewing rights for the case reports are altered at this time, so that all students can see all reports for the day. With knowledge gained from preparing a report for each of the day's cases, from the case discussion in the seminar and from reading the teaching points, the students are as well prepared as possible to perform peer review. A web page is generated which randomly links students to peer review tasks. Students are requested to have completed their reviews before midnight on the same day while the images remain available online. A graphical representation of this time line is shown in Fig. 1

The Peer Review Platform

The University of Copenhagen learning management software (Absalon, itslearning AS - P.O. Box 2686–5836 Bergen, Norway) is used via web access as the point of entry for students. Each student has a personal login to this software

Fig. 1 Time line of the peer review process



throughout their undergraduate course, and each course has its own section within the software. The imaging course page includes six folders each with links to clinical information on three cases. All students have access to the same folder each day, and they cannot preview the next day's material. A link to case information also gives access to an open text field for the student to report on the case. Since each student has an individual login, case reports are linked to the individual student author. The software requirements for this system are very basic; they are simply to present a case history and provide a text field to record input. We used the university system because it was to hand and was a familiar start point for the students, but this component could have been replaced by a more open source solution.

In addition to the above, the students were provided with URLs to a separate web server created for the course that provides five hyperlinks as follows:

1. Images and the image viewer: Students view the images via an open source medical DICOM viewer, Weasis [10] which is installed on the server. This open source viewer runs in a JAVA virtual machine (Oracle Inc, USA), on the local machine. The full set of images for the day is available for viewing through this access point only. This link is automatically updated each course day at 5 min past midnight. Images can be viewed on campus or from any location with web access. Weasis can be launched from any environment by building an XML file containing the

UID of the images to be retrieved from the server. This XML file was created using a Python script (see supplemental files) which runs on the server. The script runs each day so that the XML file is replaced with the case details appropriate for the day.

2. Comments on previous cases: This links to a summary, again hosted on the server, of pertinent findings or teaching points for the case set discussed at the previous seminar. The summary is prepared by the course teachers and is made available after the cases have been reviewed in class. This page is generated by a Python script and automatically updates at 3:00 pm each course day (see supplemental files).
3. Peer review assignments: At the start of the course, all students randomly select an ID code which allows anonymity between students in the peer review process. They are asked not to share this with their peers but the ID code is recorded against their name and student number by the teaching staff. A computer script written in Python runs on the server and randomly assigns students via their anonymous code to the cases they are to review. It checks that students are not assigned to review their own cases and that as far as possible are assigned to multiple report authors. In practice, with three cases available per day, each student will write three reports, assess three reports from three different authors, and receive three peer review assessments (one per case reported) from three different reviewers. These assignment tasks are displayed on a web page generated by the Python script each course day at 3:00 pm (see supplemental files).
4. Peer review form: This link directs the student to an online Google form, running on Google servers, that students use to enter the peer review data. The form presents five assessment criteria concerning the report. A one to 5 scale is used to indicate the presence of some desirable features in the report (1 = not at all, 5 = very much so); the higher the score, the greater the merit. The criteria used are as follows:
 1. Does the report separate findings from interpretation from comments?
 2. Does the report identify significant technical/radiographic issues?
 3. Are the radiographic findings accurate and complete?
 4. Are the conclusions based on the radiological findings?
 5. Are the comments appropriate? Do they communicate the degree of certainty of the conclusions, place them in clinical context and suggest appropriate further actions?
5. Graphical display of peer review data: Students peer review assessment data accumulates as the course progresses and is displayed using Google Chart software

running on Google servers. This is accessible to all. The chart is updated in real time as assessments are received. Student codes and scores are displayed. Student names are not displayed.

The form also required that the peer assessors state their own anonymous ID code, the case number, and the anonymous ID code of the student they are assessing.

Automation

Executable computer scripts written in the Python programming language are used to generate much of the online content. These automatically generate links to the appropriate cases for each day's activity, the up to date randomized peer review assignments, and the daily update of the teaching point summaries. The only input required to run the course comprises two CSV files, one containing the case numbers of material to be made available for each course day and the other containing the random student codes. These files are prepared and uploaded to the server once only, prior to the start of the course. The Python scripts are run daily at appropriate times on the web server as CRON jobs [11]. The only other teacher input required with respect to software is to change the student viewing rights on the Copenhagen University, Absalon software. This is done daily so that after cases are reviewed in seminar, case reports are made visible to all (to allow peer review). Prior to that, individual student reports for the day's cases are hidden from view to all but the report author and teachers. The python script files mentioned above together with Python script that was used to rename and anonymize the DICOM images, and some sample data are available for download as supplemental files. These can be inspected, used, and modified as desired. The reader is recommended to view the supplemental file "readme" in order to get full value from these files.

Student Performance as Peer Reviewers

Peer review data from all students participating in the course were analyzed. For each student, the average grade received for all reports they authored was calculated. The average of all scores awarded by each student was also calculated.

The scores obtained in the concluding summative MCQ examination were compared with those obtained in the peer review exercise. Thus, three course parameters of performance were available for each student: average peer review score received, average peer review score awarded, and the score obtained in the MCQ examination.

We tested the degree and type of correlation between the score received in peer review and that received in the MCQ examination. A positive correlation would suggest that the peer review scores have some validity. Furthermore, we tested

the correlation between score awarded in peer review and scores received in the MCQ. The outcome of this test addresses the question “do higher performing students grade more (or less) severely than lower performing students?”

We also tested for differences in peer review scores received and awarded according to overall course performance. Students with MCQ scores in the first quartile were considered low performers, those with scores in the third quartile, high performers.

Student Satisfaction with Peer Review

Student feedback on the image reporting and peer review process was assessed using an online questionnaire. This contained 12 questions (nine relating to the learning outcomes from 6 days of case review and three directly concerning the peer review process). There was also a field to allow free text for comment. The questions took the form of positive statements; the students could respond in one of five steps (1 = strongly disagree to 5 = strongly agree). Thus, a high score reflects a high level of agreement with positive comments about the preparing imaging reports and peer review.

Statistical Analysis

The statistical programming software “R”, version 3.0.2 (2013), (R foundation for Statistical Computing) was used for analysis, statistical modelling, and data plotting. Correlations between peer review scores received and given, with the summative MCQ score, were examined using Pearson’s product moment correlation coefficient, Spearman’s *rho* to test for association between paired samples. For purposes of display, a linear model was fitted to plots of average peer review score received and MCQ result. The null hypothesis was that there was no relation between these two variables and a $p < 0.05$ was considered significant.

Differences between performance in peer review (scores received and given) between low and high course performers were examined using the Welch two sample *t* test. The null hypothesis in each case was that there is no difference between the groups (low performers vs. high performers) and $p < 0.05$ was considered significant.

Results

No major technical problems were encountered. Problems were limited to the JAVA installations on the students’ machines. The server does not run a security certificate with the software so the teaching URL had to be specified in the security exceptions tab in the local JAVA control panel.

A total of 82 students provided written peer review assessments and 80 received peer review data for their work. The

difference is explained by two students who did not submit reports for the cases but did provide peer review for others. All took the summative MCQ examination where the mean score was 80.6 % (SD 7.6 %), median 81.1 % (passing grade was 60 %).

There was a statistically significant positive correlation ($p < 0.01$, Pearson correlation coefficient = 0.32) between the scores students received in peer review and those obtained in the MCQ. This relationship is shown in Fig. 2 together with a fitted linear model ($y = 6x + 10.8$).

There was no detectable correlation between the grades students awarded and the eventual grade they (the assessor) obtained in the MCQ examination, ($p = 0.14$, Pearson correlation coefficient = 0.17). These data are shown in Fig. 3.

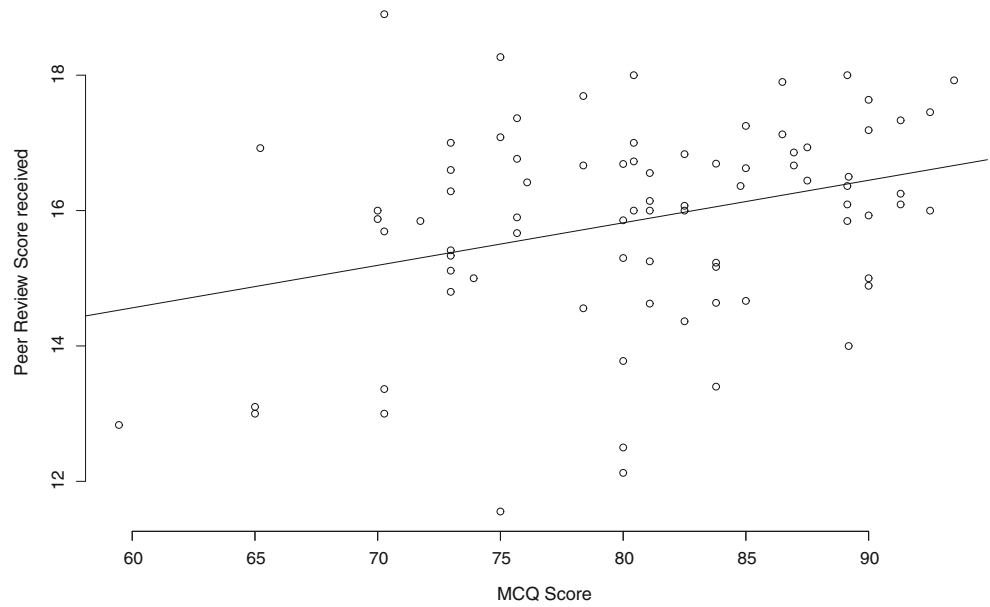
The mean summative MCQ examination score for “low performing” students was 70.1 % (SD 3.8) (passing grade was set at 60 %), while that for the “high performing” students was 90.1 % (SD 6.2). There was a statistically significant difference ($p < 0.05$) between the two groups for peer review scores received (15.3 (77 %) vs. 16.4 (82 %) respectively). A box plot showing peer review performance for both groups is shown in Fig. 4. The mean peer review score awarded by the low performing students was 15.4 (77 %) while that for the high performing students was 16.2 (81 %). This difference was not statistically significant.

Student satisfaction with the 6-day image reporting exercise was positive with a mean score of 4.19 (maximum score of 5). Their overall satisfaction with the peer review process itself was less (2.82, again out of maximum score of 5). The mean and median of the student responses to the three statements relating specifically to peer review were as follows (actual statement followed in parenthesis by average (standard deviation, (SD), and median of the students’ responses):

1. In general, I consider my peers qualified to give feedback on my work. Their comments have validity (average response 2.91 (SD 1.1) out of 5, median 2)
2. In this exercise, I learned something relevant to radiology by giving peer review (average response 3 (SD 0.8) out of 5, median 3)
3. In this exercise, I learned something relevant to radiology by receiving peer review (average response 2.55 (SD 0.9) out of 5, median 2).

Student responses in the free text format were very positive with the majority (80 %) focusing on the overall learning experiences from the cases. A minority was concerned with the peer review process itself recognizing the value in comparing reports others produced with one’s own work. Other features of the course identified to be of value by the students were the requirement to view and report in writing before dealing with the material in plenum and being “forced” to work alone for some part of the exercise. Criticisms were

Fig. 2 Relationship between scores received by students from their peers for reports they authored and the scores they (the report authors) received in the final summative multiple choice examination for the course. The line shows the best fit linear model (Pearson correlation coefficient 0.32, $p < 0.01$, for the null hypothesis that the slope of the line is zero, $n = 82$)



centered on the workload with the major complaint (20 %) being that the amount of material presented each day (3 cases) was excessive.

Discussion

The technicalities of running the course were not trivial. While the University software formed the initial point of contact for the student, extensive use of free and open source software which generated and served course web pages and images provided flexibility and control. Google forms were readily adapted for use in the collection and display of peer review data. Every course has special software requirements and

Radiology courses are no exception. Radiology is well suited to e-learning technologies [12, 13]. Experience in this course suggests that a willingness and ability on the part of both teachers and students to use a wide spectrum of software is invaluable. Ideally, enthusiasm, expertise, and ideas can come from a number of quarters, some not directly involved in the course. For this course, the provision by University Department for Blended Learning of workarounds to overcome a lack of flexibility in the in-house teaching software and by the Department of Information Technology of a virtual Linux machine with a dedicated public IP address to host the web server and python scripts was invaluable.

The result was a highly reliable supply of anonymized cases presented on a realistic DICOM viewer requiring only

Fig. 3 Relationship between the scores awarded by students to their peers for the reports they reviewed and the scores they (the assessors) received in the final summative multiple choice examination for the course ($n = 80$). No statistically significant correlation between the variables was detected using a linear model (Pearson correlation coefficient = 0.17, $p = 0.14$, for the null hypothesis that the slope of the line is zero, $n = 80$)

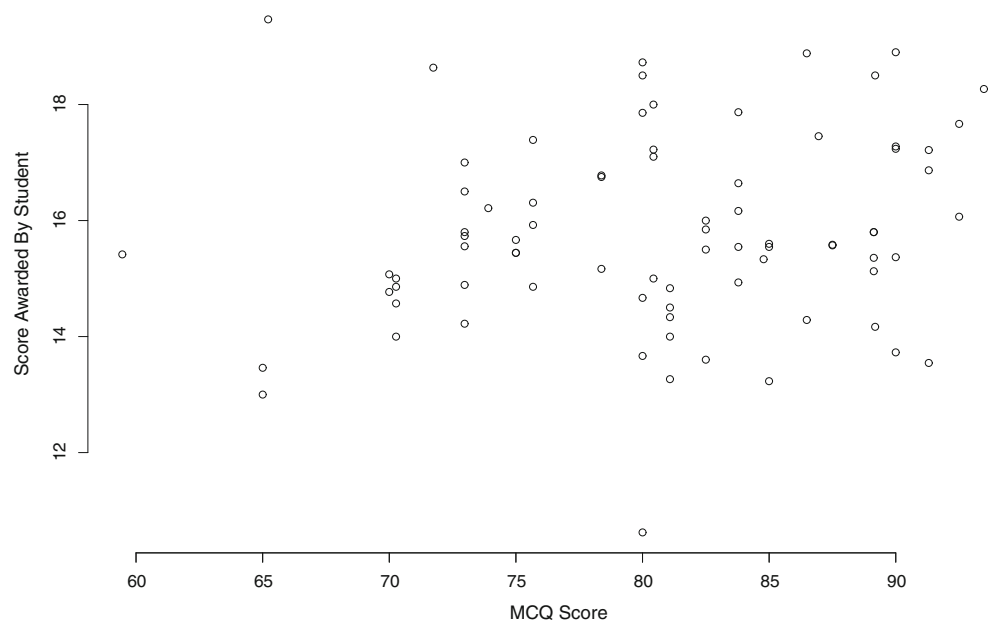
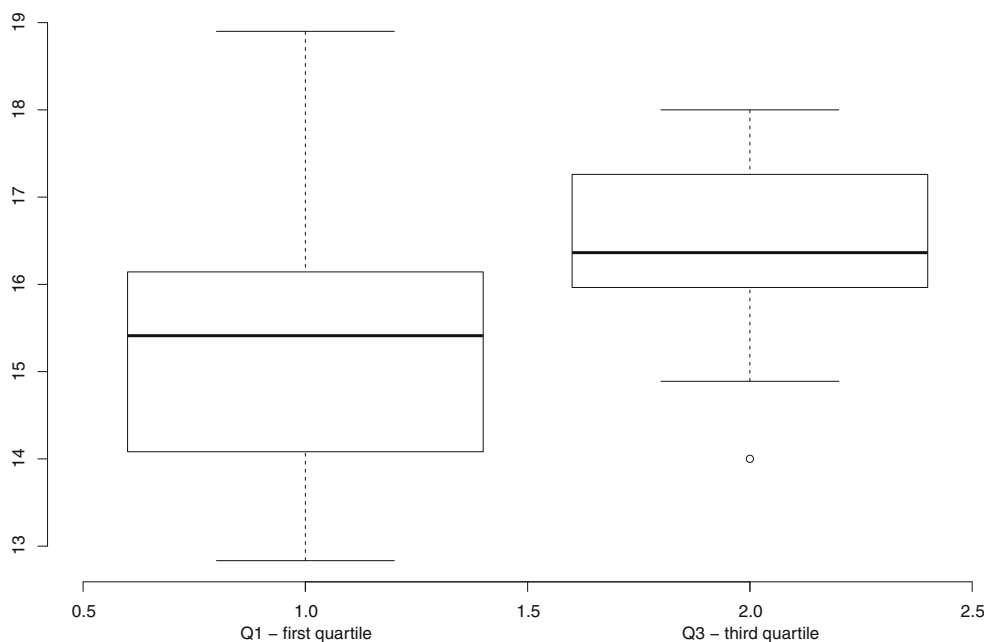


Fig. 4 Scores obtained in peer review by students grouped according to their level of course performance. “Q1—first quartile” and “Q3—third quartile” refer to students in the first and third quartiles (low and high performers respectively), when ranked according to the score obtained in the summative MCQ examination. The *box* contains the interquartile range (IQR) of the data. The *upper whisker* shows the maximum value that is less than the third quartile plus 1.5 times the IQR. The lower whisker shows the minimum value that is greater than the first quartile minus 1.5 times the IQR. One value is considered an outlier and is plotted separately ($n = 80$)



a reliable internet link and a modern browser with JAVA enabled. The platform thus could run on Windows, Mac, and Linux operating systems. The department has computers available for student use, but the students also used the university library facilities and also their own home computers. Issues relating to viewing screen quality and specification were discussed with the students as the cases progressed.

There is a weak positive correlation between the summative MCQ examination score and the scores received by peer review for the case reports. The same data were used to stratify the students into “high” and “low” course performers, and there was a difference in peer review performance between the groups. Both observations suggest that peer review has some validity; those performing the assessments can be credited with performing good reviews. That the students could perform peer review was expected, and it has been shown previously that medical students can perform peer review acceptably [14]. The correlation between peer review scores and performance with the summative MCQ examination suggests that not only were the students capable of peer review but also that the communication or case reporting task itself was relevant to student learning. The MCQ examination used for the course tests a wide range of skills: radiation safety, radiography, radiographic anatomy, disease interpretation, and knowledge of imaging modalities. While at first sight these topics may appear to span a broader range than that required to write a report on a radiological study, a deeper insight into the subject tells that the good radiologist will use knowledge and understanding in all these areas to prepare a full report. A good report will be expected to comment on the safety and radiographic issues of the study, to identify abnormalities, to interpret them, to place them in a clinical context, and to advise on

future progression for the patient. The overall aim of the Veterinary Imaging course is to enable students to produce diagnostic images safely from veterinary patients and to make valid and useful interpretations. The final 6 case reporting days are thus well aligned to the overall course and to the summative MCQ examination. The presence of a correlation can also be interpreted as an indicator that the case evaluations with peer review component of the course were well aligned with the summative examination.

The students scored the case report exercise as highly useful and relevant. However, they were less positive about the usefulness of peer review. Reasons we suggest are a perceived lack of validity and a lack of confidence. Similar student attitudes have been reported when introducing team-based learning (TBL) where peer review has an essential role [15]. The finding, however, is somewhat at odds with our previous experience of peer review, where students were more positive [9]. It is interesting that comments on the peer review process did not feature largely in the free text part of the student responses received so while the teachers considered this an important part of the course the students perhaps did not. Also, peer review is not a common technique used in teaching undergraduates at the School of Veterinary Medicine and Animal Science in Copenhagen; the exercise reported here was the first encounter of peer review for the students involved. Lesser confidence in performing peer review among junior students (second compared to third year) has been previously reported with regard to evaluation of student-generated multiple choice questions [16]. A similar effect may have played a role here. Our students were at a more junior stage in their training than the students involved in previous experiences with peer review, and in addition, the case presentation and

peer review process reported here differ substantially from that which we reported previously [9].

The impression of the teaching staff, although anecdotal and not recorded formally in the study, was nonetheless clearly positive. The 6-day case review periods had the feeling of momentum and automation; the students appeared busy and engaged and they participated fully in the afternoon sessions when cases were reviewed.

The results support our hypothesis that peer review scores have some validity and the blend of commercial software, closed but free software (Google services), and open source software Python and CronTab has proved reliable and effective. In looking forward to future courses, our plan is to continue with the infrastructure we have put in place for the delivery of this material for radiological reporting and with the peer review platform.

Acknowledgments We are indebted to Maria Thorell at the Centre for Online and Blended Learning, University of Copenhagen, for technical support, advice, and encouragement.

References

- Poe M, Lerner N, Craig J: Learning to communicate in science and engineering case studies from MIT. USA, Massachusetts Institute of Technology, 2010.
- Liu N, Carless D: Peer feedback: the learning element. *Teach High Educ* 11(3):279–290, 2006
- Brindley C, Schoffield S: Peer assessment in undergraduate programmes. *Teach High Educ* 3(1):79–87, 1998
- Biggs JB: Teaching for quality learning at university: what the student does: 2nd edition. Open University Press, United Kingdom, 2003
- Welter P, Deserno TM, Fischer B, Günther RW, Spreckelsen C: Towards case-based medical learning in radiological decision making using content-based image retrieval. *BMC Med Inform Decis Mak* 11(1):1–16, 2011
- Sluijsmans DM, Moerkerke G, Van Merriënboer JJ, Dochy FJ: Peer assessment in problem based learning. *Stud Educ Eval* 27(2):153–173, 2001
- Mahgerefteh S, Kruskal JB, Yam CS, Blachar A, Sosna J: Peer review in diagnostic radiology: current state and a vision for the future. *Radiographics* 29(5):1221–1231, 2009
- O’Keefe MM, Davis TM, Siminoski K: A workstation-integrated peer review quality assurance program: pilot study. *BMC Med Imaging* 13:19–2342, 2013
- McEvoy FJ, McEvoy PM, Svalastoga EL: Web-based teaching tool incorporating peer assessment and self-assessment: example of aligned teaching. *Am J Roentgenol* 194(1):W56–W59, 2010 [<http://www.dcm4che.org/confluence/display/WEA/Home>]
- Robbins A: *Unix in a Nutshell: A Desktop Quick Reference for System V Release 4 and Solaris 7*. O’Reilly & Associates, Inc, Sebastopol CA, USA, 1999
- Scarsbrook AF, Graham RN, Perriss RW: Radiology education: a glimpse into the future. *Clin Radiol* 61(8):640–648, 2006
- Scarsbrook AF, Graham RN, Perriss RW: The scope of educational resources for radiologists on the internet. *Clin Radiol* 60(5):524–530, 2005
- Langendyk V: Not knowing that they do not know: self-assessment accuracy of third-year medical students. *Med Educ* 40(2):173–179, 2006
- Thompson BM, Schneider VF, Haidet P, Levine RE, McMahon KK, Perkowski LC, Richards BF: Team-based learning at ten medical schools: two years later. *Med Educ* 41(3):250–257, 2007
- Rhind SM, Pettigrew GW: Peer generation of multiple-choice questions: student engagement and experiences. *J Vet Med Educ* 39(4):375–379, 2012