Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Peer-rubric evaluation in dissection courses: enhancing student performance and engagement

Joris Van de Velde
joris.vandevelde@ugent.be
Ghent University
Ali Hamdany
Ghent University
Manon Sierens
Ghent University
Senne Hellinck
Ghent University
Louis Detremmerie
Ghent University
Leander De Mol
Ghent University
Vicky Vandenbossche
Ghent University
Wouter Willaert
Ghent University

Research Article

Keywords: Dissection, Rubric, Peer evaluation

Posted Date: April 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8965440/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background:

Medical education increasingly incorporates active learning strategies, with dissection courses remaining essential for anatomical understanding. However, traditional dissection often lacks structured feedback. This study investigated the effectiveness of a peer-rubric evaluation on learning and motivation outcomes in a second-year medical course on the gastro-intestinal system.

Methods:

Students were randomized into a rubric group (n=105), who received peer evaluation based on a five-criterion rubric across seven dissection sessions, and a control group (n=309). Academic performance was assessed using a pretest, a practical posttest, and a theoretical examination. Student and evaluator experiences with the rubric were collected through surveys. Group differences in posttest and theoretical examination scores were analyzed using an independent Student's T-test. Pearson correlation analysis assessed the relationship between difference in rubric scores and academic performance. A linear mixed-model analysis was conducted to evaluate trends in rubric scores across sessions.

Results:

No statistically significant differences were found in posttest (rubric group mean 15.50 ± 2.30/20; control 15.44 ± 2.75/20, p=0.86) or theory exam scores (rubric 14.22 ± 2.75/20; control 13.87 ± 3.27/20, p=0.33). A weak positive correlation (r=0.20, p=0.04) was observed between improvement in rubric scores over time and posttest and theory performance in the rubric group. Linear mixed-model analysis showed a significant upward trend in rubric scores across sessions ($p<0.001$), indicating improved performance over time.

Survey data revealed mixed perceptions regarding peer evaluation's objectivity and learning impact. Some students valued its role in promoting engagement, while others were sceptical about fairness and influence on final scores. Evaluators found the rubric clarified expectations but noted issues with grading consistency and workload.

Conclusions:

Despite no significant effect on summative outcomes, peer-rubric evaluation appears to enhance learning progression during practical sessions. Refining rubric design, strengthening assessor training, and adjusting assessment weighting may improve its effectiveness in anatomy education.

Background

Medical education continually evolves to incorporate effective teaching methodologies that enhance student learning and practical skill acquisition (1, 2). Dissection courses are an integral part of medical training, providing hands-on experience essential for understanding human anatomy. However, traditional dissection courses often lack structured feedback mechanisms that can drive student improvement and engagement (3).

Peer evaluation has emerged as a promising educational strategy, offering structured, criteria-based feedback from peers. Studies have shown that peer evaluation can significantly improve students' critical thinking and self-regulation skills. By engaging in the evaluation process, students develop a deeper understanding of the assessment criteria and the standards for high-quality work. This reflective process helps them to internalize standards of excellence and apply these to their own work, promoting self-assessment and metacognitive skills (4, 5, 6, 7, 8).

Moreover, peer evaluation fosters a collaborative learning environment where students can learn from each other's strengths and perspectives. Research indicates that this collaborative aspect not only enhances learning but also increases their engagement and motivation (8, 9). When students know that their work will be reviewed by peers, they may put more effort into the task, resulting in higher-quality outcomes (6, 8, 9).

However, the success of peer evaluation depends on several factors, including the training provided to students on how to effectively give feedback, and the classroom culture surrounding feedback. Effective peer evaluation requires students to be objective, honest and constructive in their feedback, which can be cultivated through guided practice and the establishment of trust among peers (7). A common technique to objectify the peer evaluation process is the use of rubrics (6, 7, 10). Rubrics are structured assessment tools frequently employed for self, peer, and teacher evaluation. They typically consist of a table that outlines various criteria, with corresponding quality levels that are clearly defined by specific indicators.

The primary benefits of using rubrics are that they bring objectivity and structure to the evaluation process. Rubrics enable a straightforward and efficient approach to formative evaluation by clarifying expectations and increasing the transparency of the evaluation process. This transparency is particularly advantageous for self-assessment and peer evaluation, as it allows students to better understand their performance and areas for improvement (7, 10).

Moreover, rubrics help in standardizing the evaluation process across different evaluators, thereby reducing potential biases and promoting consistency in assessments. By clearly

articulating the criteria and quality levels, rubrics can also enhance the feedback process, making it more meaningful and actionable for students (10).

Overall, peer-rubric evaluation is a valuable tool in education, promoting active learning, critical thinking, and deeper engagement with course material. When implemented thoughtfully, it can enhance both individual and collective learning experiences in the classroom (10, 11).

This study explored the effectiveness of peer-rubric evaluation during the dissection laboratories in a second-year medical course focusing on the gastro-intestinal and endocrine systems. The aim was to determine the impact of this evaluation on academic outcomes by comparing practical posttest and theory exam scores of students who received rubric evaluations with those who did not. The impact on students' engagement was assessed with a survey about their experiences with the rubric evaluation. Finally, a short survey was conducted among the peer evaluators to gauge their experiences with the rubric system.

Methods

Participants

A total of 414 second-year medical students at Ghent University who participated in a dissection course of the gastro-intestinal and endocrine system were eligible for inclusion in this study.

Dissection Course Structure

Two weeks after a series of 21 hours of theoretical lectures on the gastro-intestinal and endocrine system, the dissection course took place, consisting of seven sessions of two hours per student spread over one month.

Evaluation Design

Both groups underwent the following assessments:

Before the practical sessions, students were required to complete a digital pretest consisting of twenty anatomy image questions, randomly selected from a database of 3,000 questions available in the learning management system of Ghent University (Brightspace®, Kitchener, ON, Canada). Students were permitted unlimited attempts over a period of 15 days. Participation in the pretest was mandatory, and the highest score achieved was recorded as the final score.

Each student had to complete seven practical dissection sessions in the dissection room. During these sessions, the anatomical regions of the gastro-intestinal and endocrine system were dissected using a dissection manual. The dissections were guided by trained anatomists and third or fourth-year medical students called tutors. The day after the last practical session, all students were required to undertake a practical posttest, conducted in the dissection room. This test was structured as a circulation assessment (scored on 20) in which dissected anatomical structures had to be identified with a time limit of one minute per structure. Additionally, two months later, students had to pass a theory exam, consisting of multiple-choice questions related to the theoretical lectures (scored on 20).

The sole intervention in the rubric group, in contrast to the control group, involved the implementation of a rubric-evaluation conducted by the tutors at the end of each of the seven practical sessions (Fig. 1).

Based on insights gained from previous years' experiences and through collaborative consultations with the teaching staff and the tutors, it was determined that five key criteria had to be evaluated in the rubric: punctuality, knowledge, communication, attitude, and execution. To establish content validity, the rubric underwent an expert review process. The rubric was created through an iterative process in which twelve students provided information on which aspects of their conduct should be evaluated, after which three experts in anatomy education provided feedback. These experts evaluated each rubric item for relevance, clarity, and representativeness. The rubric was finalized when all students and the experts agreed on its content and form. This process took 3 rounds of development.

To assess the ‘knowledge’ criterion, a list of challenging anatomical structures was created by the tutors. Three distinct quality levels were established to assess the five criteria, with scores ranging from 1 to 3, to ensure objective evaluation. In instances of student absence, a score of 0 was assigned across all criteria. A maximum score of 15 points was awarded to students who met all criteria (Fig. 2).

Immediately after each dissection session, tutors served as peer-evaluators and conducted the rubric evaluations on paper before recording them digitally. Prior to the commencement of the practical sessions, these evaluators received joint training on the proper administration of the rubric, to limit interpretation inconsistency and to enhance rater reliability. Each evaluator was responsible for assessing nine students. Students were informed in advance that the rubric scores would not count towards the final score. The evaluation was performed with the specific rubric tool in the learning management system. Students had permanent access to the rubric and the rubric scores via this learning environment, allowing them to access the feedback while preparing for the following dissection sessions.

After the posttest, a survey was conducted among the students of the rubric group about their experiences with this evaluation. A structured survey was developed for this study. The full students' survey is provided in additional file 1.

Finally, a short survey comprising 4 statements was taken from the peer evaluators to gauge their experiences with the rubric system. A structured survey was developed for this study. The full survey is provided in additional file 2.

Both surveys were conducted on a voluntary basis and were performed in the learning management system. In both surveys, a 5-point Likert scale was used (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree). Participants also had the option to give feedback in text fields.

Data analysis

An independent t-test was first performed on the pretest scores of the rubric and control groups to ensure that there was no difference in prior knowledge between them. Since students were permitted unlimited attempts until deadline, the scores were not included in additional analyses

The differences between both groups in posttest and theory exam scores were analyzed. A Levene's test was performed for the pretest, posttest and the theoretical examination scores to assess the difference in population sample variances between both groups. Then, the difference in posttest and theory exam scores between both groups were analyzed using the independent Student's T-test. To investigate whether the rubric contributed to a higher percentage of students passing, both groups were divided based on their posttest and the theory exam scores. This was achieved by creating two categorical variables (one for the theory exam and one for the posttest), with a cutoff point of 10/20 (a score of ≥ 10/20 was considered a pass). In this way, chi-square tests for 2x2 tables could be performed for both the theory exam and the posttest. This distinction was made to determine whether a statistically significant difference existed between the control group and the rubric group regarding pass rates. This approach was employed to mitigate the potential bias of medical students who consistently achieve high scores on exams. For the posttest, however, the Fisher's exact test was used instead of the chi-square test for the 2x2 table, as the expected count in more than 20% of the categories was less than 5, violating a key statistical assumption of the chi-square test.

For the rubric group, the correlation of difference in rubric scores with posttest and theory exam scores was investigated. Therefore, the variable 'difference' was created to capture the evolution of the rubric scores. The difference was calculated by subtracting the last available rubric score from the first available rubric score. A positive difference indicates an increase in scores compared to the initial rubric score, while a negative

difference indicates a decline. The correlation between this difference and the posttest and theory exam results was investigated by calculating Pearson correlation coefficient (PCC).

To investigate the learning effect of the rubric evaluation over time, rubric scores were analyzed at seven time points (T1–T7) using descriptive and non-parametric comparative statistics. The evolution of rubric scores over time was analyzed through a linear mixed model (LMM). A LMM was selected because of the presence of missing rubric scores for some individuals. This issue arose because students were permitted a maximum of two absences from the dissection courses. To avoid excluding these individuals from the analysis, their missing values were estimated using maximum likelihood estimation (MLE), enabling their inclusion in the dataset.

In interpreting the LMM, descriptive statistics were initially employed to identify trends in rubric scores over time. To formally evaluate the relationship between scores and time, a Type III Test of Fixed Effects was conducted to determine whether time had a statistically significant impact on rubric scores. Additionally, pairwise comparisons were carried out using Bonferroni-adjusted post-hoc tests to identify specific time points that were significantly different from each other. Finally, covariance parameters were estimated to quantify the residual variance, representing the portion of score variability that could not be accounted for by the time variable within the model.

Additionally, subjective feedback from the rubric evaluation survey was qualitatively analyzed to understand students' perceptions of the rubric-based assessment. Thematic analysis was employed to evaluate responses from the perception survey, identifying recurring themes, patterns, and insights across participants’ feedback.

Also, the survey of the peer evaluators was qualitatively analyzed to understand peer evaluators’ perception of the rubric-based assessment.

All statistical analyses were performed in SPSS®. The significance level was set at $p = 0.05$. The data were presented as means with standard deviations (SD) and confidence intervals (CI).

Results

A total of 414 second-year medical students participated in the study. They were randomly divided into 4 dissection groups with GraphPad randomization software (GraphPad Software®, San Diego, CA, USA). The first group was called the rubric group ($n = 105$), the other 3 groups served as the control group ($n = 309$).

The independent T-test between the rubric group and the control group on pretest scores showed no significant difference between both groups ($p = 0.051$). The estimated difference of the mean was 0.39 (95% CI [-0.78; 0.00]) less for the rubric group (Table 1).

The Levene's test showed no significant difference in population variances between both groups ($p = 0.082$).

The independent T-test between the rubric group and the control group on mean theory exam scores showed no significant difference between both groups ($p = 0.33$). The estimated difference of the mean was 0.35 (95% CI [-0.53; 0.64]) higher for the rubric group (Table 1).

The Levene's test showed no significant difference in population variances between both groups ($p = 0.073$). The independent T-test on mean post test scores revealed no significant difference between both groups ($p = 0.857$). The estimated difference of the mean was 0.06 (95% CI [-0.350; 1.044]) more for the rubric score (Table 1).

Table 1
Mean theoretical examination scores and posttest scores

	Rubric group (n = 105)	**Control group (n = 309)**	
	Mean (± SD)	**Mean (± SD)**	**P-value of the difference**
Theory exam	14.22(± 2.75)/20	13.87(± 3.27)/20	0.33
Posttest	15.50(± 2.30)/20	15.44(± 2.75)/20	0.86

To investigate whether there was an association between undergoing the rubric evaluation and passing the post-test, a Fisher's exact test was performed because not all conditions for a Chi-square test were met. No significant association was found ($p = 0.18$). The estimated percentage of passing students who underwent the rubric evaluation was 98.1% and 94.4% in the control group.

To determine whether there was an association between undergoing the rubric assessment and passing the theory exam, a Chi-square test was conducted because all test conditions were fulfilled. No significant association was found ($p = 0.13$). The estimated percentage of passing students in the rubric group was 93.3% and 88.1% in the control group.

To measure the correlation of the (difference in) rubric scores with the posttest and theory exam scores, the PCC between both variables was calculated. A PCC of 0.205 and a p-value of 0.042 were observed.

To measure the correlation of the (difference in) rubric scores with the theory exam scores, the PCC between those variables was also calculated. A PCC of 0.199 and a p-value of 0.049 were observed.

The evolution of rubric scores over time was analyzed using an LMM, revealing a clear upward trend in mean scores across the seven time points. Specifically, at T1, the mean score was 13.45/15 (SD = 1.73), while at T7, it increased to 14.6/15 (SD = 0.78) (Table 2, Fig. 3).

Table 2
Mean rubric scores by time. T; time measurements, N; number of participants, Missing; missing data, SD; standard deviation

T	N	Missing	Mean rubric score (± SD)/15
1	95	10	13.45 (± 1.73)/15
2	99	6	13.70 (± 1.32)/15
3	91	14	13.89 (± 1.60)/15
4	97	8	14.29 (± 1.30)/15
5	85	20	14.72 (± 0.73)/15
6	94	11	14.44 (± 0.86)/15
7	96	9	14.56 (± 0.78)/15

The Type III Tests of Fixed Effects shows that the effect of time is highly significant ($F(6, 65 = 13.46$ and $p < 0.001$) confirming that rubric scores change significantly over the course of the study. Pairwise comparisons of fixed effects provide additional insights into the changes at specific time points. At T1, scores were significantly lower compared to subsequent time points. For example, the comparison between T1 and T4 reveals a significant difference ($t = -3.68$, $p < 0.001$). The most substantial improvement in scores occurred between earlier time points, particularly between T1 and T4. In contrast, changes between later time points, such as T5 to T7, were smaller and non-significant ($p > 0.05$), suggesting a plateau in the improvement of scores. Covariance parameters highlight the residual variance of 1.56, which remains significant ($p < 0.001$). This indicates that, despite the overall model explains some score variability, individual differences in score variability among students continue to play a significant role.

Figure 4 illustrates the student survey on the rubric evaluation (n = 88).

A total of n = 88 students participated in the students' survey. The key findings from the students' feedback are presented in Fig. 4.

Most students agreed that the rubric evaluation positively influenced their final score or stayed neutral (72.73%; 25.00% Agree, 9.09% Strongly Agree; 38.64% Neutral). Only a minority of students disagreed or strongly disagreed with this statement (27.27%).

Similarly, a substantial portion of students felt that the peer evaluation with rubric-based grading positively influenced their learning process (70.46%; 30.68% Agree, 10,23% Strongly Agree 29.55% Neutral). Conversely, 29.54% of students disagreed or strongly disagreed with this statement.

Students generally perceived the feedback they received as valuable. A combined 78.41% of respondents agreed, strongly agreed or were neutral about the statement, "I found receiving feedback... to be valuable" (35.23% Agree, 32.95% Strongly Agree). Only 21.59% of students disagreed or strongly disagreed that the feedback was valuable.

When asked about the objectivity of the grading, students were more evenly split compared to other items. Approximately 61.37% of students agreed, strongly agreed or responded neutral to the statement that the grading based on the rubric by peer evaluators was objective (29.55% Agree, 26.14% Strongly Agree). However, a considerable proportion (38.63%) disagreed or strongly disagreed with the statement.

Lastly, a majority of 53.41% answered ‘true’ to the final ‘true or false statement’ included in the survey: “The peer evaluation project with rubric grading is a valuable addition to the practical sessions and I recommend its implementation for all groups in future years.”

Thematic analysis of open-ended comments highlighted several positive aspects, such as increased motivation from positive feedback and guidance as needed. Participants express enhanced preparation prior to the practical sessions and emphasize the importance of taking the rubric's scores into account towards the final evaluation to ensure that their efforts were accurately reflected in their grades.

Conversely, the open comments analysis revealed a few negative points, including participants' requests to have the same peer evaluator consistently, a fear of asking questions, and a perception that inappropriate behaviour remained unadjusted due to the lack of influence on their final evaluation outcomes. Additionally, concerns were raised regarding the need for clarification on point deductions, and the objectivity of the assessment is called into question.

A total of 12 evaluators (30% of total number) completed the peer evaluators’ survey. Figure 5 illustrates the results of this survey.

Regarding statement 1, which assesses clarity in understanding what was expected when grading students using the rubric (tasks, submitting forms, communicating with students),

100% of respondents agreed, or completely agreed that expectations were clear (75% Agree, 25% completely agree).

For statement 2, which evaluates clarity in understanding how to grade the students and the meaning of the different criteria in the rubric, 83.3% of respondents agreed or completely agreed (50% agree, 33.3% completely agree) and 16.7% disagreed.

In terms of statement 3, which addresses the difficulty of assessing fellow students, the responses are more mixed. Specifically, 50% disagree or completely disagreed (16.7% completely disagree, 33.3% disagree), 25% agreed or completely agreed (16.7% agree, and 8.3% completely agree) that it was difficult to assess fellow students. The reasons provided for difficulty in assessing peers include uncertainty about how strictly other peer evaluators rate and the limited time for the peer evaluator to obtain an objective and representative picture of the student's capabilities.

For statement 4, which explores whether there is a noticeable difference between students in the group graded with the rubric compared to students in other groups, 58.3% agree, 33.3% disagreed or completely disagreed.

Additionally, the survey questioned the perceived manageability of the workload for quoting students using the rubric system. In response, 75% agreed that the workload was manageable, while 25% disagreed.

Finally, peer evaluators were asked whether implementing the rubric system in all groups would be considered valuable from the perspective of the student being graded. Responses were evenly split, with 50% agreeing and 50% disagreeing.

Peer evaluators commented that students in the rubric group were generally better prepared. Many brought the necessary course materials and had a clear understanding of the dissection tasks. In contrast, those in the control group often lacked the course materials, which affected their ability to perform dissections effectively. Several respondents noted that students in the rubric group appeared more motivated, asked more specific questions, and followed the dissection plan more diligently. Open comments indicate the following opinions: some students felt that using the rubric create distance between student and evaluator and took more evaluation time than necessary, while others appreciated the rubric’s clear criteria, which helped students better understand expectations and encourage discipline. A few evaluators believed that, although there was no significant difference between groups that were graded and those that were not, the rubric could be useful in identifying students who consistently underperformed. However, some evaluators found it inconvenient to assess students using the rubric, as it requires constant monitoring, which they feel detracted from their ability to assist with dissections.

Discussion

The present study explored the effectiveness on learning and motivation outcomes of a peer-rubric evaluation in a second-year medical course focusing on the gastro-intestinal and endocrine systems. By comparing the performance of students who received rubric evaluations with those who did not, the aim was to determine the impact of this methodology on academic outcomes and student engagement.

The results indicate a noteworthy but weak positive correlation between the difference in rubric scores and students' posttest and theory exam performance. This suggests that students’ academic performance improved as they engaged more intensively with the rubric, confirming the potential of peer evaluation systems to enhance learning outcomes. This potential has also been affirmed in other studies of anatomy education (5, 12). Han et al. (5) found that students participated more actively in the dissection process, and when the peer-evaluated dissection was compared to faculty-led dissection in terms of self-assessment scores of learning objectives, scores were higher in the peer-evaluated dissection group (5). However, numerous factors could influence the observed positive correlation in this study, such as the time a student spent studying the posttest and the level of stress experienced during the exam (which was here designed as a circuit exam with a time limit of one minute per structure). This analysis provides initial insights, but further exploration is needed to understand the interplay between these contributing factors.

Despite the positive correlation between the difference in peer-rubric scores and both students' posttest and theory exam performance, the findings also reveal the absence of a statistically significant difference in overall academic performance (posttest and theoretical examination scores) between the rubric group and the control group. This outcome prompts a nuanced interpretation; although peer evaluations may foster individual learning through structured feedback (5, 12), their success appears to depend on optimizing surrounding conditions. Factors that might enhance effectiveness include comprehensive training for assessors, more objective rubric criteria, and more substantial weighting of rubric scores (10). As mentioned earlier, future analyses should also consider confounding variables such as study time and exam-related stress, as these could significantly impact posttest and theoretical examination performance (9, 13). Without these elements, the potential of peer assessments to drive considerable improvements in academic performance may remain unrealized.

To improve the objectivity and inter-rater variability of rubric design, Thomson et al. (10) propose the use of the Blooming Anatomy Tool, which can be helpful in applying Bloom’s taxonomy for assessment in a specific anatomy educational context. Valcke et al. (7) also provide concrete tips to improve the composition of rubrics: making the quality levels of a

rubric concretely observable and using qualitative terms could improve the objectivity of the rubric. Certainly at the level of the criterion ‘execution’, this could lead to a noticeable increase in objectivity. Also, using 4 quality levels instead of 3 is recommended, along with omitting the 0 criterion.

Notably, analysis of progress in rubric scores over the seven practical sessions demonstrates an improved trajectory, particularly in the initial stages. The pattern of the scores suggests an overall improvement in performance, accompanied by a reduction in variability among students as their scores stabilize. The most significant advancements were made between the early sessions, suggesting that students benefit from regular feedback as their practical skills develop. Conversely, reduced improvement in later sessions hints at a possible plateau, indicating that sustained, targeted interventions may be necessary to maintain engagement and further enhance learning outcomes over time. The observed positive progression in rubric scores may support the hypothesis that the application of a rubric, coupled with the provision of interim feedback, can facilitate the advancement of the learning process (9). Consequently, rubric-based evaluation appears particularly appropriate for formative assessment, which can play a crucial role in monitoring the development of student learning, with the primary objective of guiding and enhancing the instructional process (9).

Additionally, feedback from student surveys reflects a diverse range of experiences with rubric evaluations. While some recognized the value of peer evaluations in promoting learning and engagement, others expressed scepticism regarding objectivity and perceived lack of influence on final scores. This dichotomy in perceptions underscores the necessity for continuous reflection and adaptation of assessment practices in educational settings. Establishing a culture that values constructive feedback can help alleviate concerns regarding validity and reliability of peer assessments (6, 12).

The peer evaluators’ survey findings offer a nuanced perspective on the implementation of the rubric. While the rubric seems to provide clear expectations for most evaluators, there are areas for improvement in terms of grading clarity, the manageability of workload, and the peer evaluation process. The mixed feedback on the value of the rubric system calls for further investigation into the specific challenges facing evaluators and the potential for refining the rubric to enhance its effectivity. Future adjustments could focus on improving evaluator training, for example by introducing specific case studies during the training. Refining the rubric criteria could be possible by introducing for example specific standard knowledge questions for every student to test their anatomical knowledge. To reduce the time burden, it should be made possible to fill in the rubric directly digitally and not first on paper.

Conclusion

This study highlights the potential of peer-rubric evaluations to enhance student engagement and academic performance in a second-year human dissection course. While the results suggest a weak positive correlation between peer-rubric scores and subsequent examination performance, the lack of a statistically significant difference between the rubric group and the control group implies that further refinements are necessary to realize the full potential of this assessment method. Factors such as the quality of rubric design, assessor training, and the inclusion of more objective criteria appear critical in improving the effectiveness of peer evaluations. Additionally, the observed improvement in rubric scores over time underscores the importance of formative assessment and regular feedback in supporting student development. Despite mixed perceptions from students regarding the objectivity and impact of peer evaluations, the findings advocate for a continued evolution of rubric-based assessments to foster deeper learning and engagement. Future research should further explore the interplay of external variables like study time and stress, and work towards optimizing the peer evaluation system to maximize its educational benefits.

Abbreviations

PCC
Pearson Correlation Coefficient
LMM
Linear Mixed Model
MLE
Maximum Likelihood Estimation
SPSS®
Statistical Package for the Social Sciences
CI
Confidence Interval
SD
Standard Deviation
T1, T2, T3, ... T7
Time points (1 through 7)
n
Number of participants

Declarations

Ethics and consent to participate

This study adhered to the Declaration of Helsinki regarding ethical principles for research involving human participants. Ethical approval (ONZ-2022- 0060) was obtained from the

Ethics Committee of the University Hospital Ghent, Belgium. Informed consent was obtained electronically from all participants prior to participation.

Consent for publication

Informed consent for publication was obtained from the participants.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests

Funding

Not applicable

Clinical trial number

Not applicable

Authors' contributions

Joris Van de Velde (corresponding author submitting author): conceptualization, investigation, writing, methodology, validation, visualization, review & editing, software, formal analysis, project administration, data curation, supervision, resources

Ali Hamdany: methodology, writing, formal analysis, data curation

Manon Sierens: writing, methodology, formal analysis, data curation

Senne Hellinck: writing, methodology, formal analysis, data curation

Louis Detremmerie: writing, methodology, formal analysis, data curation

Leander De Mol: review and editing, methodology

Vicky Vandenbossche: review and editing, methodology affiliations

Wouter Willaert: methodology, review and editing, supervision, data curation

References

1. Huynh N, Burgess A, Wing L, Mellis C. Anatomy by whole body dissection as an elective: Student outcomes. J Surg Educ. 2021;78(2):492–501. 10.1016/j.jsurg.2020.07.041.
2. Antipova V, Manhal S, Siwetz M, Wimmer-Röll M, Hammer N, Fellner FA. On the added benefit of virtual anatomy for dissection-based skills. Anat Sci Educ. 2023;16(3):439–51. 10.1002/ase.2234.
3. Orsini E, Quaranta M, Mariani GA, Mongiorgi S, Cocco L, Billi AM, Manzoli L, Ratti S. Near-peer teaching in human anatomy from a tutors' perspective: An eighteen-year-old experience at the University of Bologna. Int J Environ Res Public Health. 2021;19(1):398. 10.3390/ijerph19010398.
4. Hafez SA. Design for assessment of dissection in anatomy laboratory based on group identification of structures and peer evaluation. Anat Sci Educ. 2022;15(6):1045–59. 10.1002/ase.2143.
5. Han ER, Chung EK, Nam KI. Peer-assisted learning in a gross anatomy dissection course. PLoS ONE. 2015;10(11):e0142988. 10.1371/journal.pone.0142988.
6. Nwachukwu C, Lachman N, Pawlina W. Evaluating dissection in the gross anatomy course: Correlation between quality of laboratory dissection and student outcomes. Anat Sci Educ. 2015;8:45–52. 10.1002/ase.1458.
7. Valcke M. Krachtige leeromgevingen. Gent: Academia; 2014.
8. Wisniewski B, Zierer K, Hattie J. The power of feedback revisited: A meta-analysis of educational feedback research. Front Psychol. 2020;10:3087. 10.3389/fpsyg.2019.03087.
9. Atwa HS, Potu BK, Fadel RA, Deifalla AS, Fatima A, Othman MA, Sarwani NAL, Nasr El-Din WA. Implementing formative assessment in human anatomy practical sessions: Medical students' perception and effect on final exam performance. Adv Med Educ Pract. 2024;15:551–63. 10.2147/AMEP.S465384.
10. Thompson AR, O'Loughlin VD. The Blooming Anatomy Tool (BAT): A discipline-specific rubric for utilizing Bloom's taxonomy in the design and evaluation of assessments in the anatomical sciences. Anat Sci Educ. 2015;8:493–501. 10.1002/ase.1507.
11. Hielscher AC, Everse S. Virtual peer teaching in the gross anatomy lab: A format of peer teaching and learning during the COVID-19 pandemic. MedEdPublish. 2023;12:48. 10.12688/mep.19178.2.
12. Youdas JW, Hoffarth BL, Kohlwey SR, Kramer CM, Petro JL. Peer teaching among physical therapy students during human gross anatomy: Perceptions of peer teachers and students. Anat Sci Educ. 2008;1(5):199–206. 10.1002/ase.44.
13. Bellier A, Secheresse T, Stoeckle A, Dols AM, Chaffanjon PC. Impact of background music on medical student anxiety and performance during anatomical dissections: A

cluster randomized interventional trial. Anat Sci Educ. 2020;13(4):427–35. 10.1002/ase.1918.

Figures

Figures are available in the Supplementary Files section.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfileslegends.docx
- Studentssurvey.pdf
- Peerevaluatorssurvey.pdf
- Figures.docx