Received: 24 February 2023 | Revised: 24 April 2023 | Accepted: 27 April 2023

DOI: 10.1002/ca.24058

EDUCATION

CLINICAL ANATOMY WILEY

A method for rewarding collaborative efforts in preclinical assessments

Michael F. Nolan | John P. McNamara

Department of Basic Science Education, Virginia Tech Carilion School of Medicine, Roanoke, Virginia, USA

Correspondence
John P. McNamara, Department of Basic Science Education, Virginia Tech Carilion School of Medicine, 1 Riverside Circle, Suite 202, Roanoke, VA 24016, USA.
Email: jpmcnamara@vt.edu

Abstract

Student success in basic medical science courses is typically determined by their individual performance on examinations of various types. Previous research both within and outside medical education has shown that the use of educational assessment activities can increase learning as demonstrated by performance on subsequent examinations, a phenomenon known as the testing effect. Activities primarily designed and used for assessment and evaluation purposes can also be used as teaching opportunities. We developed a method for measuring and evaluating student accomplishment in a preclinical basic science course that incorporates both individual and collaborative efforts, encourages and rewards active participation, does not compromise the reliability of the assessment outcome and is perceived by the students as helpful and valuable. The approach involved a two-part assessment activity composed of an individual examination and a small group examination with each component differentially weighted in determining an overall examination score. We found that the method was successful in encouraging collaborative efforts during the group component and provided valid measures of student grasp of the subject matter. We describe the development and implementation of the method, provide data derived from its use in a preclinical basic science course and discuss factors to be addressed when utilizing this approach to ensure fairness and reliability of the outcome. We include brief summary comments from students regarding their impressions of the value of this method.

KEYWORDS
academic performance, educational assessment, group dynamics, medical education

1 | INTRODUCTION

Student success in preclinical basic science courses is typically determined by individual student performance on written or practical examinations developed around the stated objectives of the course. For some courses examination questions are written by the faculty while for others questions are obtained from external sources such as the National Board of Medical Examiners (NBME). Advantages of the former method include the ability to directly link a particular question to a specific learning objective or to emphasize a principle or topic judged to be important. An advantage of the latter is the ability to provide students with rigorously developed and validated questions similar to those that comprise high stakes examinations such as the STEP 1 and 2 examinations.

Variability among examinations also exists regarding the types of questions developed or selected for, and included on the

examinations. Most commonly examinations are composed of multiple-choice questions (MCQ's) with or without images, tables or figures, designed to best represent the material to be learned. Specific questions may vary in difficulty from relatively simple recognition type questions to more complicated and demanding synthesis and evaluation tasks as described by Bloom and Anderson (Anderson et al., 2001; Bloom & Krathwohl, 1956). Less common, but perhaps more helpful in preparing students for the learning environment of the clinical years, are examinations composed of open-ended, short answer questions requiring students to retrieve from memory and discuss or explain previously learned information. Questions of this type are frequently structured in the oral format, a method of knowledge assessment commonly used in the clinic, operating room or other patient care settings.

An additional factor related to the assessment and evaluation processes relates to the frequency of examinations administered during a course. Some courses include multiple examinations scheduled throughout the course with each examination score contributing a specific percentage to an overall final course grade (Logan & Balota, 2008). These examinations may be limited to topics addressed during a particular unit of the course or may be cumulative in content, recognizing the progressive nature of learning in medical school. Alternatively, student success may be determined by performance on a single, summative examination administered at the end of the course. Single, end of course examinations may be perceived as more stressful in light of the amount of material that students will be responsible for and because of the relative importance placed on a single examination (Carreras & Fernandez-Castro, 1998; Dimitriev et al., 2008; Lewis et al., 2008; Vivian et al., 2003; Zeller et al., 2004).

For courses in which assessment is based on examinations with more than one component, such as those with both a written and practical/skills part, the matter of weighting of individual components must be addressed. Factors involved in these decisions include the relative amount of time allocated in the course for specific types of learning activities, resources available for learning including and the time that might be required to benefit from each. For example, anatomy examinations in which the laboratory component involves identifying particular structures on cadaveric material prepared by the students or the faculty, adequate time must be allocated for both dissection and review of the dissected specimens, particularly if all the specimens will be used in the examination. The idea of assessing student knowledge using an examination format comprised of multiple components is not new. Barremkala et al. (2020) describe an approach used for an anatomy laboratory examination which included "stations" at which students in groups of 5–6 participate in an individual skills test (IST) which involves the identification of a tagged structure and a team skills test (TST) in which students work collaboratively to respond to a patient scenario. In their model, each individual test was weighted at 50% of the overall assessment grade.

Separate from matters related to format of student assessment, that is how student learning is measured, is the issue of how student success or satisfactory performance is determined using those assessment data. For many preclinical basic science examinations, including those composed of multiple-choice questions, success is defined as an earned score that is above a particular value, often referred to as the cut score (Cizek, 1996a, 1996b; Mills et al., 1991). In some courses, the cut score is the only determinant of success with students receiving a grade of either pass or fail. In other courses, in addition to the cut score, several levels of achievement are identified such as A, B, C, or other designations intended to rank order students among those who have demonstrated satisfactory knowledge or understanding above that defined by the cut score (Njiru et al., 2020). Irrespective of the methods used for assessment and evaluation, student success in preclinical basic science courses is typically determined by a students' individual effort on an examination. Factors affecting performance on examinations are numerous, many of which incorporate the testing effect (Angel, 1949; Bangert-Drowns, Kulik, & Kulik, 1991; Bangert-Drowns, Kulik, Kulik, & Morgan, 1991a; Butler et al., 2008; Butler & Roediger, 2007; Butler & Roediger, 2008; Kang et al., 2007; McDanial et al., 2007; Smith et al., 2011) and other approaches that involve effortful retrieval and practice (Bangert-Drowns, Kulik, & Kulik, 1991; Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Butler, 2010; Karpicke & Blunt, 2011; Roediger & Karpicke, 2006).

Based on the work of previous investigators, we developed a method for measuring student accomplishment and assigning examination grades in a preclinical basic science course involving two components that would encourage and reward collaborative effort in a way that would be valid and viewed by the students as valuable. We recognized that students generally prefer active learning opportunities where they can interact with the faculty and their classmates as opposed to being passive participants in more faculty centered learning activities. We recognized also that most students also prefer to work collaboratively, especially if as a result, some meaningful benefit might accrue. We noticed that during the course, students frequently self-assess, as well as learn course material by means of practice examination questions available either commercially, created individually or passed down by previous students or classes. We noticed also that students attend and greatly appreciate review sessions offered by the faculty either as a scheduled part of the course or as an unscheduled, informal activity. The closer the review activity parallels the actual course examinations, the more favorable the activities are viewed and the more engaged the students are with them. Likewise, the more of these opportunities that are offered, the more value they seem to have.

We describe our method for assessing and valuing student performance and present examination performance data as a measure of our success in achieving these goals. We then briefly summarize student perceptions regarding this method and discuss issues to be addressed when considering the use of this method in other courses. The method described here was developed as part of an ongoing effort to encourage student engagement and participation in the course.

2 | METHODS

2.1 | Course organization

The approach described here was developed for a 90-h medical neuroanatomy course, scheduled during the final 10 weeks of the first

year of the curriculum. The course was divided into three units of instruction: 3, 3, and 4 weeks in length. Instructional methods included traditional lectures, cadaver dissection laboratory sessions focusing on the brain, spinal cord and organs of special sense and small group discussions of clinical case-based problems. Course content and organization reflected the stated course objectives which were developed by the faculty and reviewed and approved by the medical curriculum committee. The same course objectives and organization were used during both years of this report.

2.2 | Examination description

Written examinations, one for each unit written by the course faculty, were administered at the end of the 3rd, 6th and 10th week of the course, respectively. Each examination focused on content addressed during that unit with some overlap for continuity purposes. Each examination was composed of approximately 150 multiple choice questions (MCQ's), with or without images, the majority of which were structured as clinical vignettes. Examination questions were developed using a blueprint based on individual class session and course objectives, thus ensuring a balanced examination with regard to course content. Questions were structured using guidelines provided by the National Board of Medical Examiners (NBME) for the STEP examinations (Case & Swanson, 2002). Once constructed each examination was reviewed by the faculty of the course for accuracy and fidelity to the examination blueprint and by personnel from the Office of Medical Education with expertise in examination construction for validity and reliability. The same three, unit examinations were used during both years included in this report.

2.3 | Examination components and administration

Each unit examination was composed two parts: an individual component and a group component. Students were initially given 2.5 h to complete the individual component of the examination. Answers were recorded using a Scantron® answer sheet provided for this purpose. At the completion of this phase of the assessment the individual answer sheets were collected and scored by the Office of Medical Education. Performance on this part of the examination represented the individual component of the final score computed for that unit examination.

Following a short 10-min break, students were then randomized into groups of five students for the group component of the examination. Each was group instructed to take the same examination again, working collaboratively to arrive at a single group answer for each question. Groups were allocated 1.5 h to complete the group examination. New randomly assigned groups were created for each of the three, unit examinations. Students were not permitted to review study materials between the individual and group testing sessions. During the group examination students were encouraged to discuss each question within their own group and to arrive at a consensus answer for each question. Answers arrived at by the group were recorded on a single Scantron® answer sheet which was again collected and scored by the Office of Medical Education. Performance on this examination represented the group component of the final score for that examination.

2.4 | Calculating and assigning examination scores

Each student received an overall examination score based on their performance on both the individual and the group component of the examination. The individual component was represented by the percentage score earned by each student on the individual component of the examination. The group component was the percentage score earned by the group working together. All five students in a group received the same percentage score for the group component. In calculating the final examination numerical score for each student, the score obtained on the individual component was valued at 93% and the score obtained for the group component was valued at 7%. The cut score for each examination was 70%. Success on each examination and for the course was recorded as pass or fail.

2.5 | Solicitation of student comments regarding the assessment approach

Student comments regarding each preclinical course are solicited anonymously on an end of course survey administered by the Office of Medical Education. Two questions dealing with the assessment method were included on the survey: (1) *Did you find this approach helpful/valuable or not helpful/not valuable? Please comment*, and (2) *Did you feel your input was unrecognized or not valued by your group? Please comment*.

3 | RESULTS

Student performance on three examinations during the course for two consecutive academic years is presented in Table 1. Mean Individual student scores on examinations 1, 2 and 3 ranged from 80% to 84% for year 1 with similar average scores for year 2. The range of scores varied from 45% to 94% in year 1 and from 40% to 97% in year 2. Mean group scores were higher than mean individual scores for all examinations in both years ranging from 81% to 96% in year 1 and 81% to 97% in year 2 (Table 1).

The number points added to a student's individual score and the number of students who received additional percentage points for each examination using the formula for calculating final examination scores is presented in Table 2. Most students received no additional points added to their individual score. A larger number of students received either 1 or 2 percentage points with only 2 or 3 students receiving 3 additional points.

Table 3 illustrates the number of students who would have failed the examination (cut score 70%) for each examination based on their individual score and those earning a passing score as a result of

TABLE 1 Mean scores on individual and group examinations for two consecutive years.

	Year 1				Year 2			
	Individual ($N = 152$)		Group ($N = 30$)		Individual ($N = 151$)		Group ($N = 30$)	
	Mean (range)	SD	Mean (range)	SD	Mean (range)	SD	Mean (range)	SD
Exam 1	83 (45–92)	5	90 (81–95)	3	84 (40–92)	5	91 (82–96)	3
Exam 2	80 (51–94)	4	89 (83–96)	3	83 (53–94)	4	89 (81–96)	3
Exam 3	84 (52–94)	4	90 (84–96)	3	83 (56–94)	4	91 (85–97)	2

TABLE 2 Number of students benefitting from performance on group examinations.

Percentage points added based on group examination scores												
	Exam 1				Exam 2				Exam 3			
Points added	0	1	2	3	0	1	2	3	0	1	2	3
Number of students gaining points												
Year 1 ($N = 152$)	96	41	13	3	101	39	10	2	108	32	10	2
Year 2 ($N = 151$)	92	51	6	2	100	45	4	2	111	33	5	2

TABLE 3 Number of students scoring below the cut score based on individual versus group examination scores.

	Year 1		Year 2	
	Individual	Group	Individual	Group
Exam 1	5	3	6	4
Exam 2	3	1	3	2
Exam 3	4	3	4	2

factoring in performance on the group examination. Our data indicate that during the first year, 5, 3, and 4 students would have earned a failing score on examinations 1, 2, and 3, respectively, had performance been based only on their individual examination scores. However, using the method described here, the number of failing students was reduced to 3, 1 and 3 students on examinations 1, 2, and 3 respectively. A similar trend was observed for the second year with 6, 3, and 4 students posting a failing score on examinations 1, 2, and 3, respectively, based on their individual scores with only 4, 2, and 2 students failing each examination using the combined score approach.

4 | DISCUSSION

Following discussion among the faculty we decided to implement a method in which the grade a student would be assigned for each examination in our course would be determined by two differentially weighted components: an individual component and a group component. Informal discussions with upper classmen (former students of the course) regarding this idea indicated that students would likely value such a mechanism whereby a part of an examination grade, albeit a small one, would be derived from a group effort which might potentially raise a students' overall examination score. We also appreciated the likelihood that some additional learning might be achieved as a result students' working together, in this case, in an assessment activity.

4.1 | Assigning value for individual and group efforts

An important issue we had to address to ensure fairness with this method was to determine what percentage of the overall examination score should be assigned to the individual and the group components. Importantly, we understood that the score a student receives on an examination must accurately and reliably reflect that student's individual success in learning or mastering the material included on the examination. We therefore decided that knowledge demonstrated on an examination taken by the individual would and should contribute the largest percentage of an overall examination score with a smaller percentage derived from a group effort.

In determining the percentage to be assigned to each component, we considered several factors. Since each student in the group would be assigned the same group score, regardless of their participation and contribution to the group effort, we understood that care must be used to avoid giving non-participating students relatively more benefit than would be given to the more participatory or engaged students in the group. We appreciated that if the assigned percentage is too high, non-participating or minimally participating students would likely receive a relatively greater benefit than students who may not have participated to a meaningful degree. We recognized also that in some instances if the percentage was too high, students who might have scored below the cut score based on the individual score might

be assigned a passing score based on calculations using the group score. The higher the group score percentage, the greater number of students would benefit from this calculation and the criticism of failing to identify students with insufficient or inadequate understanding of the course material could be justified.

Conversely, we considered the effect this approach might have if the group percentage was too small. If the assigned percentage for the group component was set too low, the contribution of the group score would be reduced and the incentive to work as a group would be likewise reduced, an outcome we hoped to avoid. In this situation the learning benefits of group effort would also be similarly limited.

To address these concerns, we performed hypothetical calculations using different percentages for the individual and group examination scores. Based on these calculations, we determined that the individual component would be valued at 93% and the group component would be valued at 7% of the final examination score. Our belief was that these percentages would both fairly reflect a student's individual effort and provide a meaningful incentive for students to engage in a collaborative assessment effort.

Regarding the matter of time allocated to the group examination, our decision to allocate 1.5 h for the group component was based on our experience during previous years with group testing during which time group efforts were used only for providing immediate feedback regarding examination performance without any effect on the assigned examination score.

4.2 | Other effects using this assessment approach

We also considered other effects this method might bring about. For example, a student who scored a 98% on the individual component, but was a member of a group that scored 89% on the group component, would be assigned an overall score of 97%. Thus, this student would receive an overall lower score than if performance was determined by individual performance alone. This student would in effect be penalized by this method. Had the assigned percentages been for example, 88% for the individual component and 12% for the group component, the possibility of a high scoring student being penalized even more would increase. Although the possibility existed, we decided in advance that any student whose final examination score would have been reduced by this method would not be penalized by this determination. In reviewing the calculations and scores assigned, we found that no student fell into this category and thus, no student received a lower score than that based on their individual performance.

We acknowledge that determining an examination score for a particular student based on efforts that involve other students raises important concerns regarding the validity of the assigned grade and the fairness of the grade assigning process itself. If a students' score on an examination is based solely on their own individual effort, as is commonly done, little question can be raised regarding the validity of the assigned grade. Conversely, if a students' score is derived solely from a group effort and all students in the group receive the same score more significant questions regarding fairness might be raised. We found after calculating final examination scores using the formula indicated above that the number of percentage points added to any individual student ranged from zero to three, with only a few students receiving three points (Table 2). Thus, for most students, there final score represented their individual knowledge of the material tested on the examination. In light of our finding that no more than three percentage points were added to the score of any individual student, we feel confident that the calculated final examination score fairly represents the students' knowledge and understanding of the material considered on the examination.

Although raising a student's score from a failing to a passing grade was not the intent of this effort, we anticipated that the contribution of the group score might possibly result in assigning a passing score to a student who would have otherwise been given a failing score. Since the maximum number of points added was 3, this result would only affect those students who scored between 67% and 69%, or 1–3 points below the cut score. It would not benefit students who scored more than 3 points below the cut score. We found that the addition of points based on the group effort affected the score of only two students in four of the six examinations and one student in the other two examinations (Table 3). Thus, in only two instances was an unsatisfactory performance on an examination set aside.

The question may be raised regarding whether those students who would have failed the examination based on their individual effort, but who received a passing score based on the contribution of the group score, might not benefit from remediation. Our position is that the benefit of problem solving and working together in a small group setting that involves discussion among their peers is of greater overall educational value than that gained by requiring proscribed remediation effort with subsequent re-examination for those students who scored one or two percentage points below the cut score. Our belief is strengthened by the work of Butler, Roediger, Karpicke and others regarding the testing effect and its contribution to learning. (Butler, 2010; Butler et al., 2008; Butler & Roediger, 2008; Karpicke & Blunt, 2011; Roediger & Karpicke, 2006).

Our assessment approach recognizes the value of student-centered, collaborative work as a method of facilitating learning and is consistent with the views of other investigators (Agius & Stabile, 2018; Benè & Bergus, 2014; Bentley & Hill, 2009; Bruno et al., 2016; Hendelman & Boss, 1986; Manyama et al., 2016; Yeager, 1981; Youdas et al., 2008). Nonetheless, we recognize that further studies will be required to more reliably determine the extent of actual learning.

While encouraging group or collaborative effort was not a specific objective of our course, we believe that collaborative approaches to learning and assessment, like collaborative efforts in other health care delivery settings, are effective and successful. We believe assessment approaches similar to the method described here can be effectively utilized in other preclinical courses and are worthy of further evaluation.

5 | SUMMARY OF STUDENT COMMENTS

Student comments regarding the course including the examination scoring method used here were obtained as part of an anonymous, institutionally managed end of course evaluation process. Regarding this approach to grading and assigning examination scores, students were asked to respond to the following two questions: *Did you find this approach helpful/valuable or not helpful/not valuable?" Please comment* and "*Did you feel your input was unrecognized or not valued by your group?" Please comment*.

The first and most prominent theme to emerge was an appreciation of the opportunity to increase their score on each examination through a percentage derived from a group effort.

Students commented that they appreciated the immediate feedback they received from their group members, specifically mentioning the value of explanations regarding specific questions related to topics they were unclear about or were insufficiently familiar with. In addition, students recognized that the group score would likely be higher than the individual score and therefore, even if the mathematical contribution of the group score was small, it would nonetheless be valuable. Several students commented that, in addition to the possibility that point(s) might be added, that they felt they learned something in the process.

To the question: *"Did you feel your input was unrecognized or not valued by your group?" Please comment* no students responded in the affirmative. Likewise, no comments were offered indicating that one or more individuals in the group dominated the discussions or decision making despite the fact that, some students were more knowledgeable better prepared than others.

Several students commented, based on their feelings immediately following the individual component of the examination, that they felt anxious about their performance and that the discussion and feedback gained during the group component lessened some of their anxiety. A few students who believed that they did poorly or possibly had failed the examination commented that they felt less comforted by the feedback they obtained. These students typically contacted the course director shortly after the examination about possible causes for poor performance and to discuss strategies for improvement in the future.

Unsolicited comments included requests for more frequent opportunities to answer "practice questions" in a group setting with faculty feedback, suggesting that this would help them identify topics where their knowledge and understanding was weak well before the end of unit examination. Two students commented that their apprehension of poor performance was corroborated during the group examination and recommended that the approach be discontinued.

6 | LIMITATIONS

We recognize that our approach is based on an examination format that included only a written component and may not be applicable for courses that include laboratory or practical assessments. We also appreciate that our curricular structure allowed for three, four-hour long testing sessions, including breaks, an allocation of time that may not be feasible in other curricula.

We acknowledge also that the percentages we assigned to each component of the examination represents our collective judgment based on discussions and calculations using hypothetical examination scores. It is possible that for courses of different class sizes, with formative and/or only summative examinations, with examinations that have different mean and range scores or that have a different number of examinations, that the percentages we assigned might not be considered as appropriate or justifiable as they were for us.

7 | CONCLUSIONS

The method described here was designed to encourage and reward interaction and cooperation among small groups of students in an assessment activity in a preclinical basic science course. The method involved a written examination composed of two parts: an individual examination and a small group examination. Each component contributed a specific percentage to the overall examination grade with the largest percentage being derived from performance on the individual examination. We found that the approach facilitated active engagement and collaboration among students without compromising the validity of the assessment data. We believe the percentages assigned for each part were adequate to facilitate cooperative work while not compromising our ability to reliably determine individual student success. Additionally, the method provided students with prompt feedback regarding their likely performance on an examination. For most students, this feedback was supportive regarding their learning of the material. For those less confident, the group discussions offered an opportunity to correct areas of incomplete or inadequate understanding. While we recognize that all members of the group may not have contributed equally, the narrative comments from the students suggest that none felt that they were not given the opportunity to contribute nor were they excluded from participating in the process. We believe that our efforts to encourage and reward collaborative, student-centered learning have been successful and are worthy of further study.

FUNDING INFORMATION

No funding was received for conducting this study. Authors have no financial or non-financial interests to disclose.

ORCID

Michael F. Nolan https://orcid.org/0000-0002-9177-4276
John P. McNamara https://orcid.org/0000-0002-5178-7750

REFERENCES

Agius, A., & Stabile, I. (2018). Undergraduate peer assisted learning tutors' performance in summative anatomy examinations: A pilot study. *International Journal of Medical Education*, *9*, 93–98. https://doi.org/10.5116/ijme.5aa3.e2a6

Anderson, L. W., Krathwohl, D. R., & editors. (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. Allyn & Bacon.

Angel, I. G. W. (1949). The effect of immediate knowledge of quiz results on final examination scores in freshman chemistry. *The Journal of Educational Research*, *42*, 391–394. https://doi.org/10.1080/00220671.1949.10881704

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*, 213–238. https://doi.org/10.3102/00346543061002213

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. C. (1991). Effects of frequent classroom testing. *The Journal of Educational Research*, *85*, 89–99. https://doi.org/10.1080/00220671.1991.10702818

Barremkala, M., Taylor, T. A. H., & Taranikanti, V. (2020). Team assessment in the laboratory setting (TAILS): A novel approach using cadavers to assess collaborative learning in the gross anatomy lab. *Medical Science Educator*, *30*, 21–22. https://doi.org/10.1007/s40670-019-00860-9

Benè, K. L., & Bergus, G. (2014). When learners become teachers: A review of peer teaching in medical student education. *Family Medicine*, *46*, 783–787.

Bentley, B. S., & Hill, R. V. (2009). Objective and subjective assessment of reciprocal peer teaching in medical gross anatomy laboratory. *Anatomical Sciences Education*, *2*(4), 143–149. https://doi.org/10.1002/ase.96

Bloom, B. S., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: the classification of educational goals, by a committee of college and university examiners. In *Handbook I: Cognitive domain*. Longmans Green.

Bruno, P. A., Love-Green, J. K., Illerbrun, S. L., Holness, D. A., Illerbrun, S. J., Haus, K. A., Poirier, S. M., & Sveinson, K. L. (2016). Students helping students: Evaluating a pilot program of peer teaching for an undergraduate course in human anatomy. *Anatomical Sciences Education*, *9*, 132–142. https://doi.org/10.1002/ase.1543

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *3*, 1118–1133. https://doi.org/10.1037/a0019902

Butler, A. C., Karpicke, J. D., & Roediger, H. L. I. I. I. (2008). Correcting a meta-cognitive error: Feedback enhances retention of low confidence correct answers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 918–928. https://doi.org/10.1037/0278-7393.34.4.918

Butler, A. C., & Roediger, H. I. I. I. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*, 514–527. https://doi.org/10.1080/09541440701326097

Butler, A. C., & Roediger, H. L. I. I. I. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory and Cognition*, *36*, 604–616. https://doi.org/10.3758/MC.36.3.604

Carreras, P., & Fernandez-Castro, J. (1998). Adaption to examination stress in first and second year university students. *Studia Psychologica*, *40*, 181–187.

Case, S. M., & Swanson, D. B. (2002). *Constructing written test questions for the basic and clinical sciences* (3rd ed.). National Board of Medical Examiners.

Cizek, G. J. (1996a). Setting passing scores. *Educational Measurement: Issues and Practice*, *15*(2), 20–31.

Cizek, G. J. (1996b). Standard-setting guidelines. *Educational Measurement: Issues and Practice*, *15*(1), 12–21.

Dimitriev, D. A., Dimetriev, A. D., Karpenko, Y. D., & Saperova, E. V. (2008). Influence of examination stress and psychoemotional characteristics of the blood pressure on heart rate regulation of female students. *Human Physiology*, *34*, 617–624. https://doi.org/10.1134/S0362119708050101

Hendelman, W. J., & Boss, M. J. (1986). Reciprocal peer teaching by medical students in the gross anatomy laboratory. *Medical Education*, *61*(8), 674–680. https://doi.org/10.1097/00001888-198608000-00007

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. I. I. I. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*, 528–558. https://doi.org/10.1080/09541440601056620

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*, 772–775. https://doi.org/10.1126/science.1199327

Lewis, R. S., Nikolovam, A., Chang, D. J., & Weekes, N. Y. (2008). Examination stress and components of working memory. *Stress*, *11*, 108–114. https://doi.org/10.1080/10253890701535160

Logan, J. M., & Balota, D. A. (2008). Expanded vs. equal interval spaced retrieval practice: Exploring different schedules of spacing and retention interval in younger and older adults. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, *15*, 257–280. https://doi.org/10.1080/13825580701322171

Manyama, M., Stafford, R., Mazyala, E., Lukanima, A., Magele, N., Benson, R., Kidenya, E. K., Msuya, S., & Kauki, J. (2016). Improving gross anatomy learning using reciprocal peer teaching. *BMC Medical Education*, *22*(16), 95. https://doi.org/10.1186/s12909-016-0617-1

McDanial, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, *14*, 200–206. https://doi.org/10.3758/bf03194052

Mills, C. N., Melican, G. J., & Ahluwalia, N. T. (1991). Defining minimal competence. *Educational Measurement: Issues and Practice.*, *10*(2), 7–10. https://doi.org/10.1111/j.1745-3992.1991.tb00186.x

Njiru, S. M., Karuku, S., & Nyaga, M. (2020). Performance rankings in education: Implications for policy and practice. *Educational Research Review*, *15*(12), 700–710. https://doi.org/10.5897/ERR2020.4071

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Smith, M. K., Wood, W. B., Krauter, K., & Knight, J. K. (2011). Combining peer discussion with instructor explanation increases student learning from in-class concept questions. *CBE-Life Sciences Education*, *10*, 55–63. https://doi.org/10.1187/cbe.10-08-0101

Vivian, N. G., Koh, D., & Chia, S. E. (2003). Examination stress, salivary cortisol and academic performance. *Psychological Reports*, *93*, 1133–1134. https://doi.org/10.2466/pr0.2003.93.3f.1133

Yeager, V. L. (1981). Peer teaching in gross anatomy. *Medical Education*, *56*(11), 922. https://doi.org/10.1097/00001888-198111000-00010

Youdas, J. W., Hoffarth, B. L., Kohlwey, S. R., Kramer, C. M., & Petro, J. L. (2008). Peer teaching among physical therapy students during human gross anatomy: Perceptions of peer teachers and students. *Anatomical Sciences Education*, *1*, 199–206. https://doi.org/10.1002/ase.44

Zeller, A., Handschin, D., Gyr, N., Martina, B., & Battegay, E. (2004). Blood pressure and heart rate of students undergoing a medical licensure examination. *Blood Pressure*, *1*, 20–24. https://doi.org/10.1080/08037050310025645

How to cite this article: Nolan, M. F., & McNamara, J. P. (2023). A method for rewarding collaborative efforts in preclinical assessments. *Clinical Anatomy*, 1–7. https://doi.org/10.1002/ca.24058