

Using Peer Assessment and Feedback to Support Individual Learning

Jason Hall, University of Manchester, United Kingdom

The Asian Conference on Education & International Development 2019
Official Conference Proceedings

Abstract

Background: Feedback supports learning but providing detailed individual feedback is time consuming. Involving students in peer marking and providing constructive feedback can enhance student engagement. Delegating marking and feedback has the potential to save staff time but inter-rater variability limits the value. Higher levels of reliability are obtained when markers just decide which of two assignments '*is best*'. This project employed a series of adaptive comparative judgements (ACJ) to overcome the inter-rater variability.

Method: Students were assigned ten pairs of assignments and for each pair they judged which was best. An algorithm used this series of multiple comparative judgements (A is better than B, B is better than C etc.) to create a rank order. Students were asked to provide constructive feedback on each assignment reviewed. Staff reviewed the appropriateness of the student feedback and moderated the rank order before using it to assign individual marks to assignments.

Results: 149 students submitted assignments. 143 students completed the peer review component making 1,415 comparative judgments. The rank order generated by ACJ was found to be in broad agreement with staff judgements during the moderation process. Each assignment received feedback from 6-10 students. The mean length of feedback was 350 words per assignment (range 50-500 words). The length of feedback was not related to the rank order.

Conclusion: A series of comparative judgements can be used to address inter-rater variability in peer marking. Further work is required to explore the effectiveness of peer generated feedback.

Keywords: Peer assessment, Adaptive comparative judgment, Peer feedback

iafor

The International Academic Forum

www.iafor.org

Introduction

Assessing student assignments is obviously a vital part of education. It enables students to appreciate how their work has been judged and it is also used summatively for progression or to obtain an award. It is frequently necessary to employ several assessors to assess an assignment with large cohorts of students and achieving consistency between assessors can be difficult. It is known that some assessors have a tendency to be ‘hawk’ like whilst others are ‘dove’ like (McManus et al 2006).

Involving students in the assessment process can enhance student engagement and increase the student’s exposure to different approaches to an assignment which could help them to reflect more deeply on their own approach to the assignment (Morris 2001). Delegating the assessment to students has the potential to save staff time but inter-rater variability limits the value (McManus et al 2006).

Deciding which mark to award an assignment can be a challenge for staff. A student would obviously feel aggrieved if they felt their work was assessed by a *hawk* whilst their colleague’s work was assessed by a *dove*. One way of addressing this difficulty involves comparing two pieces of work and deciding which is best. Higher levels of reliability are obtained when markers just decide which of two assignments ‘*is best*’ (Pollitt 2012). If these comparisons of pairs are repeated many times then they can be used to create a rank order. Such comparisons can be conducted manually or using computer software. A manual assessment could involve comparing each piece of work with every other piece of work and awarding the winner of each comparison one mark. The total number of marks each piece of work receives can be used to create the rank order and this can, with scaling, be used to generate a percentage mark if the first and last piece of work is marked by the member of staff (Hall 2018).

The provision of written feedback along with the grade for an assignment can support student learning. Students frequently demand greater levels of feedback to support their learning but providing detailed individual feedback is time consuming. Involving students in providing constructive feedback is an attractive option because it has the potential to further enhance student engagement and to help to address student demand for feedback (Morris 2001). In addition, a mismatch between how students and staff describe helpful feedback has been reported (Blair et al 2012). Involving students in the feedback process has the potential to provide the type of feedback that students want.

This project aimed to use comparative judgments to overcome inter-rater variability of peer marking of a reflective assignment. It employed software called CompareAssess to conduct a series of adaptive comparative judgements (ACJ) to create a rank order. The project also used the software to allow students to provide feedback to their colleagues.

Method

All final year pharmacy students were required to submit a written assignment which included a reflective account based upon a workshop they attended. The reflective account was limited to a maximum of 500 words and was submitted online via the virtual learning environment (VLE). The assignment brief provided students with list

of the assessment criteria which were grouped under the following headings; presentation and language, reflection on learning from workshop; reflection on potential impact on future role as pharmacist; and future learning objectives.

An eLearning technologist uploaded all the student assignments onto the ACJ software called CompareAssess. This programme then assigned each student with ten pairs of assignments. Students were presented with one pair of assignments at a time and for each pair they judged which was best (A is better than B, A is better than C, D is better than A etc.). Students were provided with a guide to comparing the pairs of assignments which was based upon the assignment brief (see figure 1).

Criteria	Comparison
Presentation and language	Which assignment: <ul style="list-style-type: none"> demonstrated the best use of English throughout? was the easiest to understand? had the most typos or grammatical errors? was best presented? (e.g. clear headings, spacing, consistent font etc)
Reflection on learning from workshop	Which assignment: <ul style="list-style-type: none"> included the best discussion of what the student has learned from the workshop? demonstrated the best understanding of the topic?
Reflection on potential impact on future role as pharmacist	Which assignment: <ul style="list-style-type: none"> had the best discussion of how their future practice might be influenced by this topic?
Future learning objectives	Which assignment: <ul style="list-style-type: none"> included the best description of what the student would like to learn more about or experience in the future to develop their understanding of the topic? included the most reasonable and feasible plans?
Figure 1: Student guide for comparing pairs of assignments	

An algorithm within CompareAssess used this series of multiple comparative judgements to create a rank order. The rank order was moderated by staff before it was used it to assign individual marks to assignments. Marks were awarded to the nearest 5%. It was not necessary to grade each individual assignment but rather staff reviewed a sample of assignments to identify the cut off point for each of the 5% grade boundaries. For example; Student 1 = 90%; Students 2 to 10 = 85%; Students 11 to 24 = 80%; Students 25 to 44 = 75%; students 45 to 70 = 70% etc.

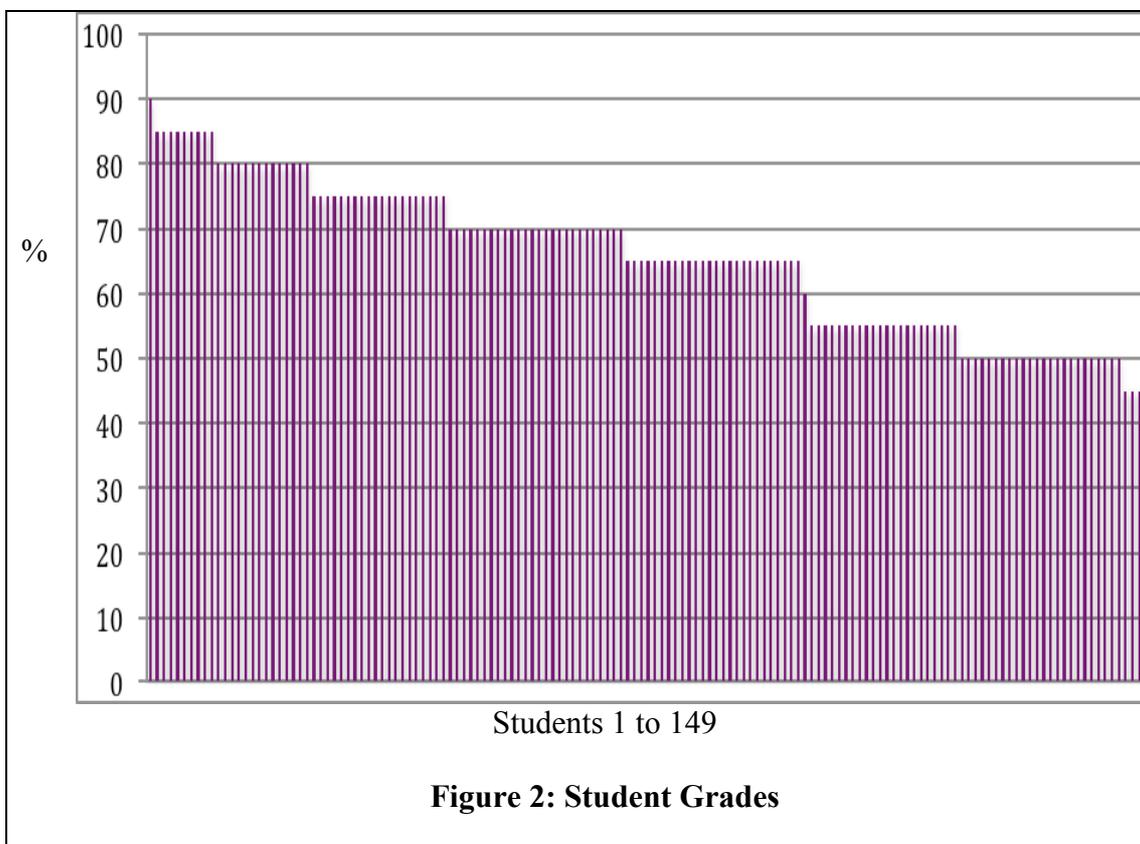
Students were also asked to provide constructive feedback on each individual assignment they reviewed. Staff reviewed the content of the student feedback to ensure there were no offensive statements but they did not make any judgement on the appropriateness of the comments for individual assignments.

Results

All students on the cohort (n=149) submitted a reflective account. The majority of students (n=143) completed the peer review component making a grand total of 1,415 comparative judgments. The rank order generated by adaptive comparative judgements was found to be in broad agreement with staff judgements during the moderation process.

(A) Grading

Staff reviewed the first and the last reflective account from the rank order and assigned marks of 90% and 45% respectively. Staff then reviewed a sample of reflective accounts in between in order to find the cut off point for each 5% boundary. The grades awarded can be seen in figure 2.



(B) Feedback

Each assignment received feedback from 6-10 students. The mean length of feedback was 350 words per assignment (range 50-500 words). The length of feedback was not related to the rank order. No attempt was made to assess how helpful the feedback was to students. Examples of student feedback can be seen in figure 3.

Student 1

“This is a well written and comprehensive reflective account. The structure was easy to follow and you have thought in-depth about how the learning applies to your future practice. I particularly liked how you applied current and past learning/skills to the situations, such as consultation skills. To improve you could comment on the method of learning (i.e. the workshop) and whether you believe it was affective for you.”

Student 2

“A decent log that explains what was learnt well however it doesn't particularly reflect on the workshop very well”

Student 3

“Good but didn't really explain your feelings about the session. Could have applied it directly to your learning/future practice more”

Student 4

“Good description of learning but action plan isn't well developed; explains what you want to learn but not how you will do so.”

Student 5

“Good description of what the student has learnt from the workshop. The plan described is also very realistic and sensible”

Student 6

“There was probably a little too much information included relating to the particulars of the learning in the workshop - e.g. results of opinions of your group and of the task. This account is more about reflection, not telling the reader what you did step-by-step.

Excellent impact on future career - really showed how you learned how the pharmacist can fit into all of this.”

Figure 3: Examples of student comments.

Discussion

Adaptive comparative judgments can be used to generate a rank order and this can be used to generate individual marks for students. Staff are not required to assess each individual assignment and therefore there is a potential time saving for academic staff. Such work does require input from an eLearning technologist and further work is required to ascertain the total amount of staff time (academic and eLearning technologist) required to complete the assessment. Educators must also consider the license costs of the software. Alternatively, manual methods could be used to make the comparative judgments and this would eliminate the license costs but the practicalities of issuing random assignments to students as well as collecting and collating all the results must be considered.

Adaptive comparative judgments helped to overcome inter-rater variability, which otherwise is likely to have been problematic with so many students assessing work. The marks obtained from this assessment were sufficiently robust to enable them to

be used summatively. This is particularly helpful if peer marking is being used to increase student's exposure to different pieces of work and to produce a grade required for progression. Whilst this work did not consider whether adaptive comparative judgements can overcome the challenges posed when staff hawks and doves are required to mark assignment within a large cohort, it would seem likely that it should be able to cope with several staff markers if it can cope with over 140 different students.

A relatively large volume of comments was generated for each assignment. The comments contained a mixture of positive feedback (what they liked) and constructive feedback (what they thought was needed to improve the work). However, it is not possible to comment on the value of this feedback as this was not investigated.

Conclusion

A series of comparative judgements can be used to address inter-rater variability in peer marking. Further work is required to explore the effectiveness of peer-generated feedback.

References

Blair A, Curtis S, Goodwin M, Sheils S. What feedback do students want? *POLITICS: 2013 VOL 33(1)*, 66–79

Hall J. It's not fair: hawks and doves. *Pharmacy Education*, 2018; 18 (1) 249 - 250

McManus IC, Thompson M, Mollon J. Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP(UK) clinical examination (PACES) using multi-facet Rasch modelling. *BMC Medical Education*. 2006 **6**: 42

Morris J. Peer assessment: a missing link between teaching and learning? A review of the literature. *Nurse Education Today*. 2001 21(7):507-13; 513-5

Pollitt A. The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy and Practice*. 2012 **19** (3): 281-300

Contact email: Jason.hall@manchester.ac.uk