

Feedback from peer and teacher assessments of introductory anatomy essays



Priti Pandey

The University of New South Wales, Sydney, Australia
p.pandey@unsw.edu.au

Doug Magin

The University of New South Wales, Sydney, Australia
d.magin@unsw.edu.au

Abstract: *The study investigates the content and character of qualitative feedback given by peer and teacher assessors to students who submitted a general essay assignment in Introductory Anatomy. The analysis consisted of classification of peer and teacher comments on 102 students' essays. The basis of classification was the identification of which criterion categories were alluded to in the general comments on each feedback sheet. With one exception, the content and character of peer assessment comments were very similar to those made by teachers. The exception was the incidence of failure to comment. Almost one-quarter of student assessments contained no written comment, whereas all teacher assessments contained written comment. A major finding from the study was the large variation between teachers in the criteria they focused their comments on, and the difficulty this poses for students in developing an understanding of what constitutes successful essay writing.*

Keywords: *teacher and peer assessment, essays, feedback.*

Introduction

In a recent article, Holroyd (2000) advocated the need for academics to adopt a more professional approach to their assessment practices. The areas identified for change included (i) increased emphasis on learning enhancement, (ii) increased attention to formative rather than summative aspects, (iii) more frequent provision of descriptive comment and constructive feedback, and (iv) less reliance on assessment by teaching staff alone and more involvement of self, peer and workplace assessors (p. 29). These concerns have been raised in other studies. Mowl and Pain (1995) have argued that '...the assessment process must generate meaningful formative feedback for the learner, and not just a mark or grade' (p. 326). According to Stefani (1998), the assessment feedback most useful to the student is that which provides the performer with direct useable insights into his/her current performance, yet this is often not realised:

It is still the case that too many academics believe that a grade, and a short series of comments, usually ... simple praise or blame ... constitute feedback, when what students actually want is user friendly information, relating to how they are doing and how specifically they might be able to improve upon what they are doing (Stefani, 1998, p. 348).

Similar needs for change have been made with respect to peer assessment practices. Topping et al. (2000) have urged an increased emphasis on formative assessment which ‘...yields rich and detailed qualitative feedback information about strengths and weaknesses, not merely a quantitative mark or grade’ (p. 150), and have alluded to the dearth of case studies concerning the nature and quality of the feedback provided by peers:

Almost all existing studies of the reliability and validity of peer assessment utilise comparison of marks or holistic grades, rather than of more open-ended, qualitative, formative feedback. ...There is an evident need for more reliability and validity studies of purely qualitative peer assessment’ (Topping et al., 2000, p. 152).

Peer assessment is no longer a novel concept, and was seen by Brown and Dove (1991), over a decade ago, as becoming a ‘mainstream idea’. What is surprising is the lack of higher education literature on peer assessment of essays when essays continue to be a dominant feature of university assessment methods. Topping et al. (2000) is apparently the only study, which systematically analyses and compares feedback comments made by peers and teachers on essays.

Comparing peer and teacher feedback

Recent reviews of peer assessment (Topping, 1998; Dochy et al., 1999) have supported claims on the learning benefits of involving students in the process of assessing another students’ work, and of giving and receiving feedback. They document numerous studies where the experience of peer assessment resulted in the development of better understanding of the learning task, and of the criteria used to judge their work. Educators have also claimed longer-term benefits relating to the development of professional autonomy and judgement.

Despite a consensus on the potential educational benefits of involving students in assessment, little is known about the character of the comments made by students in assessing essays, or the extent to which qualitative feedback from peers differs from that made by teachers. The study by Topping et al. (2000) appears to be the first, which has attempted such an investigation. Their study involved a cohort of 12 students undertaking a postgraduate educational psychology course in Scotland. Students were required to write a report of approximately 5000 words on a topic relating to child development or analysis of intervention techniques. The students used a feedback protocol sheet listing 15 criteria including ‘clear conceptualisation of main issues’, ‘literature review’, ‘critical awareness’, ‘precision of language’, and ‘conclusion/synthesis’ and provided a purely formative assessment of one other student’s essay. They were instructed to ‘...make at least one qualitative comment under each criteria heading’, and to ‘...prefix comments with + (indicating comment on aspect adding to the value of the work), – (aspect detracting value from the work), 0 (neutral comment)’ (p. 156). The same feedback protocol was used by the two teaching staff who provided both the formative feedback assessment and an overall summative mark. The data were then analysed for similarity of semantic content of feedback statements made by each teacher and peer assessor for the same essay.

The study found that, overall, staff and peer assessments showed a very similar balance between positive and negative statements, indicating good agreement in identifying strengths and weaknesses of the essay being assessed. However, when comments were analyzed for semantic similarity, ‘...only half of all formative assessment statements made showed some degree of correspondence between staff and peers’ (p. 163).

Background to the Study

Students enrolled in Introductory Anatomy at the University of New South Wales undertake a number of assessment tasks, including an end of session exam. One of the assessable tasks is the 'general education assignment', constituting 10% of the final mark. The students choose any one from a list of ten topics and write a 1000-1200-word essay. In previous years the assignments were marked by the teachers who conducted tutorials in the subject, most were part-time tutors. In 1999 the first author introduced peer assessment in addition to teacher assessment for this assignment, with peer and teacher assessment contributing to 5% marks each.

Prior to undertaking the essay assignment, tutors discussed with students the introduction of peer assessment and the criteria for assessing the essays. The feedback and assessment protocol sheet contained the marking criteria and procedures to be followed by students and tutors. It consisted of two parts: a score on a 5-point rating scale (from 'inadequate' to 'outstanding') for each of six specified criteria; and a 'general comments' section. The assessors were advised that general comments should include constructive feedback on how to improve the essay in relation to the criteria. Summative assessment consisted of a single global mark (maximum of 10) to be entered on the assessment sheet. Assessors made their overall judgements of the essay on the basis of the following criteria:

- (i) Addressing the topic
- (ii) Evidence of research on the topic
- (iii) Adequacy of discussion
- (iv) Coherence and readability
- (v) Conclusion/justification
- (vi) Referencing

Students were instructed to submit the assignments using their student numbers as sole identification, which were collected by the five tutorial staff in their tutorial classes. To ensure fairness (and anonymity) in the marking process, students in each tutorial class were assigned an essay from a different tutorial group to mark on the attached feedback and assessment sheet. After one week the tutors collected all the marked assignments from the students in tutorial classes. The peer assessment sheets were separated and retained (by the first author) and the teacher assessment sheets were attached to each essay and distributed to the five tutors. Each tutor then marked the assignments of the students from their own tutorial group. The completed feedback and assessment sheets were then photocopied for later analysis, and the essays together with the feedback sheets were returned to the students.

The current study

The study reported here has been prompted by the need, as outlined in the literature, for case-based studies into the content and character of qualitative feedback on students' written work. The analysis undertaken has been restricted to the general comments section only, involving inspection and classification of peer and teacher comments on 102 students' essays. The basis of classification was the identification of criterion categories alluded to in the general comments on each feedback sheet. In many cases the classification appeared to be straightforward and unambiguous. For example, the following feedback made by a teacher was easily classified as consisting of remarks which correspond to three of the six listed criterion categories:

- More research needed to be done. There was inadequate referencing of material. Coherence and readability was satisfactory, taking into consideration this may be the person's second language.

The second example, by a peer assessor:

- Did not adequately address the ethical issues involved. Simplistic view on the subject – you did not argue your views satisfactorily.

Here, the criterion 'addressing the topic' is explicitly referred to. However, the remaining comments required judgement before being considered to relate more closely to the 'adequacy of discussion' criterion category than to any of the other listed categories.

Results

Table 1 displays comparisons between peers and teachers with respect to the number (N) of assessments in which comments have been made in each of the listed criterion categories, and the relative frequency (%) of these in each category. The data displayed shows a reasonably close correspondence between the relative frequency of comments for the six listed criteria. The 'coherence and readability', the most frequently commented criterion category (peer/tutor ratio 30/37%) attracted approximately one-third of all feedback from both peers and teachers. The least frequently commented criterion (peer/tutor ratio 9/10%) for both groups was 'addressing the topic'. There was, however, a marked difference in the tendency to make comments in categories other than the six listed criteria. For example 9 students commented on the essay being 'interesting', and 5 students commented on the possible plagiarism by incorporating or copying material available on the Web. None of the tutors commented on this latter aspect.

Table 1: Content analysis of comments by peers and teachers

<i>Comment Category</i>	<u>Peers</u>		<u>Teachers</u>	
	N	%	N	%
Addressing the topic	12	9	19	10
Evidence of research	12	9	19	10
Adequacy of discussion	20	14	22	11
Coherence and readability	42	30	73	37
Conclusion/justification	14	10	22	11
Referencing	19	13	34	17
Other	17	12	8	4
<i>Totals</i>				
Total discrete comments	141		197	
Number essays assessed	102		102	
Comments per assessment	1.4		1.9	

As can be noted peer markers commented on fewer categories (1.4) than did teacher assessors (1.9). However, this average includes 25 peer feedback assessments, which contained *no comments at all*. For the 77 peer markers who did offer comments, the average number of categories they commented on was almost identical to that made by teaching staff. Word length of comments made was also found to be very similar – an average of 22 words for peer assessors, as against an average word length of 20 for teacher assessors.

The relative frequency of comments on criterion categories by peer assessors was distributed broadly similar over each of the five tutorial groups. However, the profiles for the five teacher assessors were found to be strikingly different as displayed in table 2 (N = number of assessments in which comments were made in each of category). For example, 92% of all

comments by tutor 5 related to just two criterion categories, namely ‘conclusion/justification’ and ‘referencing’. By contrast, 71% of comments made by tutor 2 related to just one criterion – namely ‘readability’.

Table 2: Profile of comment categories for individual teachers

	Tutor 1	Tutor 2	Tutor 3	Tutor 4	Tutor 5	Total
	N	N	N	N	N	N
Addressing the topic	4	2	5	8	-	19
Evidence of research	9	3	4	3	-	19
Adequacy of discussion	3	2	8	7	2	22
Coherence and readability	20	20	18	15	-	73
Conclusion/justification	2	-	2	6	12	22
Referencing	6	1	5	12	10	34
Other	-	-	3	5	-	8
Total discrete comments	44	28	45	56	24	197
Number assessed	18	19	22	25	18	102
Comments per assessment	2.4	1.5	2.0	2.2	1.3	1.9

Teacher feedback

Each of the five teachers who assessed a group can be distinguished as being distinctly different in choosing the criterion category to comment on. The character of these differences is indicated by illustrative quotes for each of the five teachers.

Teacher 1: This teacher provided 66% of comments in two areas ‘readability’ and ‘research’. As illustrated in the quotes, comments often follow a similar groove:

- A well researched paper and passionately written
- A well researched paper. A little work is needed on writing style, which will allow the paper to flow. As it is, it seems to ‘jump’ a little.

About one-third of comments, however, referred to other criteria – as, for example:

- Addressed the topic very well by defining and classifying out-of-body experiences. Another reference or two would have been good.

Teacher 2: A majority of comments (71%) related to ‘readability’, often focused on grammar and terminology:

- The fire chief a ‘*charred area*’? Well written – use more commas to punctuate ideas
- Some sentences require additional punctuation to make them easier to read and interpret
- Avoid starting too many sentences with ‘And’, ‘But’. Avoid ‘we get’. Are ‘dissections’ really the same sort of things as ‘experiments’?

Teacher 3: Here, the focus of most (57%) of the comments related to two categories – ‘readability’ and ‘discussion’.

- Well written but arguments are repetitive and not convincing. A broader focus is necessary to argue the point better. Alternatives to animals in research should have been presented more forcefully
- Very interesting points mentioned, but the explanations to all such points are inadequate. I doubt the student fully understands what is written in the essay.

On a number of occasions, other criterion categories are included in this teacher’s comments:

- A better structure could have been built around the presented ideas. Good work but broader discussion is necessary. What about the bibliography?

Teacher 4: This teacher provided a more varied set of comments, although almost half (48%) were classified as relating to two areas – ‘readability’ and ‘referencing’. The character of this teacher’s comments are illustrated in the following quotes:

- Very well written, argued and supported by references. Congratulations. It is outstanding!
- It is a good, well researched and written essay, except for in-text referencing. It is hard for the reader to work out the source of information if it is not cited in the text

This teacher often referred to other criterion categories, as exemplified in the next two quotes:

- The topic on ‘your reaction to seeing cadaveric material’ based on past experience, was quite OK- but the ethical issues were not handled adequately
- Well argued and discussed essay. You have justified your arguments reasonably.

Teacher 5: This was the most extreme of the five in that almost all (92%) of the comments made related to just two categories – ‘referencing’ and ‘conclusion/justification’. The focus of the comments on the latter category was that of urging the student to come to a statement of their own judgement or opinion on the topic they investigated:

- You *must* reference your essay. On what basis do you make your conclusion?
- I want *your* opinion on whether the phenomenon exists!
- Referencing needs to be better. What is your *opinion*?
- Good – your own opinions rather than hypothetical questions would be better in the conclusion.

Peer feedback

There were no detectable differences between the five groups in terms of the relative frequency of peer comments on the different criterion categories. The majority of student feedback comments had a quite similar character to those made by teaching staff:

- Good discussion. Although you have completed the research, you need to do in body referencing. A few slight fluency problems. But a good essay!
- I thought that some of the points presented were vague and too general, and needed more details for clarity of understanding. Proof-reading? Lots of grammatical errors. Conclusion wasn’t very strong. There is also evidence of a clash of language ability between introduction and conclusion with body
- Weak arguments, not enough supporting evidence given. Talked about irrelevant issues (computer aided teaching). Also, it is unnecessary to talk about animal rights.

As indicated in table 1, students were much more likely to mention aspects unrelated to the six specified criteria. Nine peer assessors referred to ‘interest’ as an indicator of essay merit. For example:

- It was good how you limited your discussion of the topic to two specific areas rather than attempting to cover everything in the 1,000 or so word limit. You created an interesting introduction to the realm of outer-body-experience. Cool!
- Well done – a really interesting essay.

Another aspect, also mentioned by five students (but not by teacher assessors) related to possible plagiarism:

- Seems to be inconsistent in the language used – is it your own work? Excellent essay except for the parts which were downloaded. It would have been better to write in your own style rather than someone else’s

- Inadequate referencing may indicate that parts were ‘pasted’ off the net, however, I have not checked it. In addition, several grammar mistakes which I would not expect a uni student to make.

Finally, there were a number of instances in which a student’s essay received teacher feedback (and mark) which were sharply at odds with that received from the peer assessor. The first quote, from a staff member was accompanied by a mark of 8 for the essay.

- Good – your own opinions rather than hypothetical questions would be better in the conclusion.

This same essay, however, received a mark of 5 from the peer assessor, with quite distinctly critical comments:

- References could have been better. Should be more scientific than ‘unsolved mysteries’ and ‘strange and unexplained mysteries’. Grammar could have been better, as well, the general organization of discussion into readable and ordered paragraphs. The essay was written at the last minute as errors could have been fixed if reread.

Character of comments

The large majority of the comments made by assessors (75% of peer assessments, and 78% of teacher assessments) were found to have included constructive comments on how the essay could be improved. In those cases where general comments did *not* include constructive feedback, the tenor of the comment was usually one of praise:

- Excellent– a very coherent and well-researched essay. Well done!

From the analysis of comments made by students, only 6 instances were found where criticisms were made, and for which no constructive feedback was given to address the shortcomings:

- Should an essay start off with such a detailed example?

A similar result was found for teacher comments. Here, there were 5 comments containing criticisms, which were not accompanied by specific feedback on how to improve the essay:

- Strong discussion and research, unclear presentation of case though.

The major difference between peer and teacher assessments related to the number of instances in which no comments at all were advanced. All teacher assessments contained general comments, whereas comment was absent in 24% of peer assessments.

Discussion and concluding remarks

The overall impression gained from the study was that the feedback comments from peers and teachers were similar with respect to content and character. One area of difference was that students were much more likely to make evaluative comments on areas other than the six specified categories, and particularly in relation to ‘interest’ and ‘plagiarism’. The most worrying aspect of peer feedback was the number of peer assessments, which contained no general comments at all. Future efforts will be directed at persuading students of the importance of providing constructive written feedback.

A major finding, which we had not anticipated, was the striking difference between the five teachers with respect to the criterion categories to which their comments alluded. It is presumed that the aspects which staff focus most of their comments on are a manifestation of the criteria by which they judge the merit of an essay (Topping et al., 2000). If this were so,

we would come to the conclusion that these staff held very different (if unstated) criteria for judging essays. An alternative explanation for the wide divergence on what teachers comment on, is that the pressure of time in marking a large number of essays results in the practice of using formulaic responses. What is commented on, in essence, is what quickly comes to hand as useful feedback for the student writer. This would imply that there is no necessary nexus between what aspects are commented on, and the criteria which are used to judge a student's work (Magin, Helmore & Baker, 2001). Whatever the reason for these large differences, it is clear that students will be hindered in developing insight into how their written work will be judged by teachers - even within the same course. If students perceive that there is a wide variation between teachers in what is expected of them in essays, then confusion can set in (Lea & Street, 1998).

We had set out to compare the character and content of qualitative feedback on students' written work by peer and teacher assessors, anticipating that this would enable us to judge whether peer feedback on essays would be helpful in yielding '...rich and detailed qualitative feedback information about strengths and weaknesses' (Topping et al., 2000). On this count it was seen as generally successful. What has emerged from the study is the need to focus attention on teacher feedback practices. The experience of peer assessment has the potential to assist in developing students' understanding of what constitutes successful essay writing, and of being able to judge their own and the work of others. But antecedent to fulfilling that potential, ways will need to be found to ensure that teacher feedback provides a clear and consistent model on which students can base their understanding.

References

- Brown, S., & Dove, P. (1991). *Self and peer assessment*. Birmingham: SCED.
- Dochy, F., Segera, M., & Sluijsmans, D. (1999). Review of self, peer and co-assessment in higher education: a review. *Studies in Higher Education*, 24, 331-350.
- Holroyd, C. (2000). Are assessors professional?: student assessment and the professionalism of academics. *Active Learning in Higher Education*, 1, 28-44.
- Lea, M., & Street, B. (1998). Student writing in higher education: an academic literacies approach. *Studies in Higher Education*, 23, 157 - 172.
- Magin, D., Helmore, P., & Baker, J. (2001). Assessing students' oral communication skills- which skills? In Z. Pudlowski (Ed.), *Proceedings of the 4th UICEE Annual Conference on Engineering Education* (pp. 237-241). Bangkok: UICEE.
- Mowl, G., & Pain, R. (1995). Using self and peer assessment to improve students' essay writing: a case study from geography. *Innovations in Education and Training International*, 32, 324-335.
- Stefani, L. (1998). Assessment in partnership with learners. *Assessment and Evaluation in Higher Education*, 23, 339-350.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249-276.
- Topping, K., Smith, E., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment and Evaluation in Higher Education*, 25, 149-169.

Copyright © 2002 P. Pandey & D. Magin: The authors assign to HERDSA and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to HERDSA to publish this document in full on the World Wide Web (prime sites and mirrors) on CD-ROM and in printed form within the HERDSA 2002 conference proceedings. Any other usage is prohibited without the express permission of the authors.