Routledge
Taylor & Francis Group

ARTICLE

Double-blind multiple peer reviews to change students' reading behaviour and help them develop their writing skills

Jan Ketil Rød [ID][a] and Marte Nubdal[b]

[a]Department of Geography, Norwegian University of Science and Technology (NTNU), Trondheim, Norway; [b]Faculty of Social and Educational Sciences, NTNU, Trondheim, Norway

ABSTRACT

For a geography bachelor course about climate change, we replaced the end-of-course exam with one term paper and three term-paper peer reviews. Our objectives were to design a learning environment where students read continuously throughout the semester, develop their writing skills, become familiar with quality criteria for academic texts, and get trained in applying these. To support students in their term-paper writing and term-paper peer reviews, we arranged two annotated-bibliography exercises as optional learning activities. A t-test demonstrated a statistically significant increase in performance for those who participated in these exercises compared to those who did not. A survey confirmed that students still doubt their own and their peer students' capability to provide authoritative reviews, but qualitative interviews supported the findings that a majority of students found the peer-review process valuable for their reading behaviours and the development of their writing skills. The improvements, however, were mostly related to form (such as structure, grammar, and how to set up a proper reference list) and less related to academic content.

Introduction

Many university professors, especially those in the humanities and social sciences, hold firmly to the belief that reading compliance is integral to learning (Hoeft, 2012). The ability to search for relevant research articles, to read these although they may not be part of the required course reading, and comprehend what one is reading are skills that academics typically would want their students to achieve (Smith, 1982). Students who not only comply with the required reading but also seek out additional materials to enhance their learning are advanced learners (Clump et al., 2004). Unfortunately, the amount of reading that students do has been steadily declining for the past several decades (Burchfield & Sappington, 2000), and most students do not complete all required reading assignments and cram just before the exam (Lei et al., 2010). Consequently, there are few advanced learners. Students who do not read, either because of a lack of motivation (Guthrie & Alao, 1997) or because of poor reading comprehension (Ryan, 2006), come to

---

CONTACT Jan Ketil Rød ✉ jan.rod@ntnu.no 🖃 Department of Geography, Norwegian University of Science and Technology, Trondheim NO-7491, Norway

rely exclusively on other learning material (e.g., lectures) and might therefore not reach their potential learning outcome.

Next to reading, many university teachers consider writing skills central to learning in their disciplines (Badcock et al., 2010). Academics often complain about students' poor writing skills (Lillis & Turner, 2001), but the development of such skills is conducted in an ad hoc manner, and students are expected to learn academic writing without explicit instruction (Taffs & Holt, 2013). Explicit training in academic writing does not seem to be normal practice in higher education,[1] or examples of such provisions are under-documented in the literature (Ferguson, 2009). Students need support to master academic writing (Wingate & Tribble, 2012). "If being a good writer is paramount to a student's future academic or professional success, practicing and crafting creative and reflective writing must be considered a fundamental component of higher education" (Burlingame, 2019, p. 59).

This article is about providing extrinsic incentives for students to read continuously throughout the semester and to design learning environments where they can develop their writing skills. In the sections below, we describe the course where we have implemented changes, the learning interventions we carried out, and how we implemented double-blind multiple peer reviews. Thereafter, we provide theoretical arguments for using assessment to change study behaviour and grounds for using peer review arranged as formative assessment. As educational support for writing term papers and term-paper peer reviews, we gave students annotated-bibliography exercises, and we describe the content of these. Our assumption was that students benefited from the first of these optional exercises when writing their own term paper. We test this assumption for statistical significance. We report results from a questionnaire about reading behaviour among students and their perceptions of the review process and from semi-structured interviews about their perceptions of the implemented interventions and how this affected their reading behaviour and the development of their writing skills. Finally, we discuss our findings and provide some concluding remarks where we sum up the results from the three research aims this article pursue:

(1) Examine whether the annotated-bibliography exercises enhanced student performance.
(2) Explore students' attitudes and perceptions of the implemented interventions on assignments and whether they asserted a positive influence on (a) their reading behaviour and (b) the development of their writing skills.
(3) Explore how students considered the reviews from their peers.

Context

Bachelor course on climate-change effects

The context for this study was a course on the effects of climate change as part of a bachelor geography programme. The curriculum included the physical basis for climate change, the effects of climate change with a special emphasis on the possible increase in frequency and intensity of extreme-weather-related events, as well as issues related to communicating climate-change effects. Climate-change adaptation was a core focus of

the course, but various mitigation options were also part of the curriculum such as the REDD+ initiative to combat deforestation and forest degradation in the tropics. We gave the course in English, and there were up to 75 students registered for the course each year; about half of them were international students. We have offered this course since 2007. At the beginning, the course was based on rather traditional teaching and assessment methods (lectures, self-study, and final exam). After running the course for four years, we included a term paper as a compulsory task that students needed to pass to take the final exam. On the basis of another five years of experience, we concluded that introducing a compulsory term paper did not change the students' reading behaviour nor did it increase the students' writing skills.

From 2017, we replaced the single assessment based on a final exam with three kinds of assessments with weights as given in brackets: (1) results from nine mini-exams held during the first half of the semester automated by our Learning Management System (30%), (2) a term paper (40%), and (3) three term-paper peer reviews (30%). We filmed the lectures and clipped them into shorter knowledge clips of 10 to 15 minutes, and as such, we transformed nine traditional lectures into nine learning modules with six or more knowledge clips. Designing a flipped and digital learning environment enabled us to move resources previously spent on lecturing to other learning activities aimed to support students working on their term paper and term-paper peer reviews.

Inspired by Mulder et al. (2014), we allowed students to choose one topic from four available topics (Mulder et al. used five). The topics were about (A) sea-level rise and the Maldives, (B) climate refugees, (C) deforestation, and (D) extreme-weather events. In addition to writing their own term paper, students needed to write term-paper peer reviews for three students writing about the three other topics. Because each student needed to go deep into each of the topics, as a term-paper author on one topic and as a reviewer for the three other topics, we assumed it should not matter too much which topic a student chose. We shortened the list of assigned reading but adapted it so that the four term-paper topics were covered by parts of the assigned reading. We assumed that reading for a term paper and for term-paper peer reviews would be more motivating than reading for an end-of-semester exam, and we treated writing as an active learning process through which students learned about how to relate to and communicate their ideas with their peer students.

Students had about five weeks to write the first version of their term paper. Once students had submitted their term paper, we organised the peer-review process. For about 25 days, the students worked on their reviews. Finally, students had about three weeks before the deadline for the submission of their final term paper. As support for students writing their term paper and term-paper peer reviews, we arranged learning activities on literature search, use of reference-management software, and annotated-bibliography exercises.

Annotated-bibliography exercises

An annotated bibliography is a list of citations of books, articles, and documents. Each citation is followed by a brief descriptive and evaluative paragraph, the annotation (Olin Library Reference, 2018). Like an abstract, the annotation is a descriptive summary of the academic work, but an annotation is also a critical evaluation of the text,

including its relevance for a topic and a description of the author's point of view and his/her authority on the topic. The task for the first annotated-bibliography exercise was to search for relevant research literature that the students could use for their own term paper. The task for the second one was to search for research literature relevant for the three reviews that they needed to give. Both exercises were elective. The second annotated exercise was much more demanding than the first one, because the students needed to base their annotated bibliographies on three articles instead of one. An essential part of the term-paper requirements was to refer to text both from the assigned reading and from relevant research articles, for instance, articles found during the first annotated-bibliography exercise. An essential part of the term-paper peer reviews was to suggest relevant research articles they found during the second annotated-bibliography exercise (see Appendix 1). Thus, the annotated-bibliography exercises aimed to stimulate students to become advanced learners seeking out additional reading material beyond the list of assigned reading. The term paper and the peer reviews were extrinsic incentives for students to become advanced learners because we required students to search and find relevant articles both for their own term paper and for their three term-paper peer reviews.

Theoretical background

Using assignment to reorient students' reading behaviour

According to Gibbs (1999), assessment is a powerful lever that tutors can use to reorient the way students respond to a course and behave as learners. "A lecture may inspire a student to read more", but more influential, Gibbs assumes, is the nature of assignments and assessment criteria (Gibbs, 1999, p. 42). On the basis of experiences from Great Britain, Gibbs (2006) recognises that students are increasingly becoming strategic and use their time and efforts only on tasks that are assessed. Assessment, therefore, is an excellent way of getting students to spend time on a given task and thus a strategic measure that teachers can use to make students spend time on particular tasks (Gibbs, 2006).

Courses based on traditional lectures with no assessment during the semester, only self-study toward the final exam, tend to stimulate an unwanted reading behaviour where very little reading is done during the semester but a lot of reading is done during the days prior to the exam (Dysthe & Engelsen, 2011). Students following such learning approaches may invest only a minimal amount of effort during the semester but make a large effort during the days just before their exams. Although they may succeed in their exam results, they will probably soon forget what they have learned. When knowledge is rapidly gained, it is often also rapidly lost (Gynnild, 2003). Gibbs claims that writing a term paper stimulates a qualitative different kind of reading because students need to read more about a given topic to develop an argument (Gibbs, 1999). However, the students' term paper may easily end up being "mediocre, regurgitative, and uninspired" (Cohen & Spencer, 1993) if the students' conception of the term-paper assignment is a call for "all about" writing rather than analysis and argument (Bean, 2011). In order to promote engagement and deep learning, students need to have several writing assignments, but more important than quantity is the quality of how these are designed.

Instructors can design short scaffolding assignments early in the course to teach the skills needed to write a term paper due at the end of a semester (Bean, 2011).

Peer review as assessment

The first principle of good feedback principle reads: "helps clarify what good performance is (goals, criteria, expected standards)" (Nicol & Macfarlane-Dick, 2006, p. 205). An assessment grid or rubric may be helpful for students to clarify what is expected of a piece of work (O'Donovan et al., 2001). "Rubric design is arguably the most important aspect of a peer-review activity, since here the instructor decides both what feedbacks will be useful to authors, as well as the nature of the critical skills to be fostered in the reviewer" (Purchase & Hamer, 2018, p. 1150). It is further important to engage students in identifying quality standards applicable to their own work and in making judgements about how their work relates to these standards (Boud, 1995). Gibbs (1999) elaborates on this phenomenon, which he calls internalisation of criteria for quality. From their experience from writing, submitting, and reviewing research articles, academics understand quality standards for academic texts. Academics have internalised what the threshold standards consist of and are reasonably good at judging when an article text is good enough for publication. Students, in contrast, often hand in work, which they have barely read through quickly. "They have no idea of the standards required, and even if they did, it would not have occurred to them to apply this standard to their own work" (Gibbs, 1999, p. 47).

One way to facilitate an internalisation of quality criteria among students is to provide them with opportunities to evaluate and provide feedback on each other's work. Such peer processes help students to develop the skills needed to make objective judgements against standards, skills that are transferred when they turn to producing and regulating their own work (Boud et al., 1999). Not all students may be able to provide very good and elaborated reviews to their peers, but although peer students' feedback is of a lower quality than that from a professor, systematic use of peer assessment can provide students with more and quicker feedback. Rapid feedback is recognised as more influential to learning than perfect but late feedback (Gibbs & Simpson, 2005), and this constitutes one reason for using peer assessment. Another reason is that peer assessment may solve practical problems a tutor may have when he/she needs to provide quick feedback to many students. Peer assessment alleviates an intensive workload for instructors. As a result, there are numerous articles on the validity and reliability of peer assessment (Cho et al., 2006) and on ways to calibrate these (Balfour, 2013). Calibrating peer assessment has been relevant for justifying the use of peer assessment for summative assessment, particularly helpful when grading courses with a large number of students, such as a MOOC. Peer summative assessment is not allowed in Norway. Our motivation for using peer assessment is that it may facilitate deep learning (Sitthiworachart & Joy, 2008) and assist the development of students' writing and evaluating skills (Tsai & Liang, 2009).

Methodology

Computer-assisted multiple peer review

Our university had recently evaluated several electronic Learning Management System and had chosen Blackboard to be operational by autumn 2017. One of the reasons why our university chose Blackboard was its capability to facilitate peer assessment. Unfortunately, we realised that the peer assessment that our version of Blackboard could facilitate only worked for pair of students switching assignments for peer review. Our design for multiple and double-blind peer review was too complex for the purchased version of Blackboard. Elsewhere, academics resolve similar complex peer review by using additional software such as PeerMark, part of the Turnitin software (Søndergaard & Mulder, 2012), or PeerGrade (Wind et al., 2018). As an alternative to purchasing new software, we used TeamSite. TeamSite enables students to co-operate, work together on a text (co-writing), and share documents and other files by using Microsoft Office software. Besides, TeamSite is a freely available software for us and for our students and accessible because one can enter TeamSite from Blackboard's top menu.

We used TeamSite to administer the peer-review process, including setting up a list of all assignments for each of the four term-paper topics. Because TeamSite's activity log tagged contributors, which jeopardised the double-blind review process, we arranged TeamSite to produce anonymised templates, stored in the cloud, for both submitted term papers and term-paper peer reviews. After the term-paper submission deadline, we assigned term-paper peer reviews for all students having submitted a term paper. The number of students that could sign up for any of the four topics was restricted to a maximum of 25% of the total number of students enrolled in the course. It is rather difficult to know the exact number of students registered for a topic, and the ones who are registered do not necessarily finish the courses that they start. To allow each student to receive and give three reviews, some flexibility in how to handle unequal number of students per topic was needed (e.g., students review term-papers from previous year). The distribution of term-paper-review tasks for a subject with about 70 students is a complex operation if done manually. Fortunately, the workflow functionality implemented in TeamSite enabled us to automate the peer-review process, requiring only some manual management at certain points in the process.

Students had a little more than three weeks to finish their reviews, which they did by using peer-review templates with two parts (Appendix 1). The first part consisted of structured questions related to the three elements of the required structure of the paper (introduction, analysis and discussion, and conclusion) as well as a fourth element about the overall writing quality and organisation of the paper, such as complete and proper references. The second part consisted of a scoring rubric. Scoring done by students was not taken into consideration when we assessed the final term papers, but we used the same scoring rubric. Students knew that the assessment criteria in the template and the rubric they used for term-paper peer reviews were the same as those we would use for grading. We hoped that this made the students engage with the assessment criteria for the term papers they were reviewing in a manner that would benefit their own work (Hounsell et al., 2008).

Comparing means

Comparing the means of results obtained by one group of students who chose to use a new educational method with those from another group of students not using this method is a common way to assess whether the innovation tried out for the first group has any effect on performance (e.g., Mui et al., 2015). Similarly, we wanted to test whether the learning activities we provided students with had any effect on their learning outcome. Specifically, we were interested in the effect of the annotated-bibliography exercises.

Survey

Towards the end of the semester, we arranged a survey where the participating students evaluated the course with specific questions related to their reading behaviour and the peer-review process. We posed questions related to the students' reading behaviour (see Table 1) and their perceptions of the peer-review process (see Table 2). Among the 14 questions, 10 were closed questions, and 4 were open questions. The responses were anonymous, and students participated voluntarily. We received responses from 49 students (out of 67), leading to a response rate of a little more than 73%. We formulated the closed questions as statements and asked students to rank their agreement by using scores on a Likert scale with five categories. Tables 1 and 2 show how the responses were distributed among the five categories, and whenever a statement has

Table 1. Survey results related to reading behaviour (numbers are percentages). 2 = strongly agree, 1 = agree, 0 = neither agree nor disagree, -1 = disagree, -2 = strongly disagree.

		2	1	0	-1	-2	N.A.
1	The term-paper assignment was an important incentive for me to read the assigned readings.	**14.29**	**46.94**	12.24	16.33	8.16	2.04
2	The term-paper assignment was an important incentive for me to search for recent research articles that I could use for my term paper.	**53.06**	40.82	2.04	4.08	0	0
3	Reading literature for the term-paper assignment resulted in better learning than if the reading had been for a traditional end-of-semester exam.	**18.37**	**42.86**	26.53	10.20	2.04	0
4	Reading literature for the three review assignments resulted in better learning than if the reading had been for a traditional end-of-semester exam.	4.08	26.53	26.53	30.61	8.17	4.08

Table 2. Survey results related to perceptions of review process (numbers are percentages). 2 = strongly agree, 1 = agree, 0 = neither agree nor disagree, -1 = disagree, -2 = strongly disagree.

		2	1	0	-1	-2	N.A.
5	The reviews from my peer students helped me improve my term paper (overall **helpfulness**).	**20.41**	**59.18**	10.20	6.12	2.04	2.04
6	The reviews from my peer students were **balanced**; they highlighted both strengths and areas for improvement.	**14.29**	**57.14**	18.37	6.12	2.04	2.04
7	The reviews from my peer students were **insightful**; the reviewers pointed out things I had not thought of myself.	**10.20**	**57.14**	16.33	10.20	2.04	4.08
8	The reviews from my peer students were **helpful**; they contained specific suggestions that I was able to implement.	**18.37**	**61.22**	12.25	4.08	0	4.08
9	The reviews from my peer students were **authoritative**; the reviewers had read the literature and knew their stuff.	2.04	38.78	38.78	14.29	2.04	4.08
10	The reviews from my peer students were **clear**; the reviews were well written and easy to understand.	**14.26**	**61.22**	16.33	4.08	0	4.08

either a majority of responses on the positive or the negative side of the scale, it is marked with bold font.

Semi-structural interviews

To further understand the effects of interventions and whether these provided students with extrinsic incentives to read and help develop their writing skills, we carried out eight semi-structural interviews. The interviews were carried out in accordance with standard ethical procedures: the students volunteered to take part, offering informed consent. They were aware that they could withdraw from the interview at any time or ask us to delete the empirical data resulting from their participation later, with no consequence and without having to explain their decisions. Interviewees were also assured that their comments would be reported anonymously. Four of the informants took the course in 2017 and the other four took the course in 2018. Among the eight individuals, there were five females and three males. The interviews were structured around five topics: reading habits, quality criteria for academic texts, writing, annotated-bibliography exercise, and peer-review process. The interviews took about 30 minutes each and were recorded with the permission of the interviewees, transcribed, and analysed.

Results

The effect of having performed annotated-literature exercises

All grades are letter grades, but we converted these to numbers (A = 5, B = 4, C = 3, D = 2, and E = 1). Out of those having their term-paper assignments approved (two failed), 29 out of 65 students submitted the first annotated exercise. We call this group of students ABE 1, and we call the group of students who did not do this exercise ABE 0. The group of students handing in the first annotated-bibliography exercise had an average grade of 3.62, which was significantly higher[2] ($p < 0.05$) than that of the group of students who did not do this exercise, having an average grade of 3.08 (see Figure 1).

　　The means of students' grades from their term paper (as shown in Figure 1) seem to support the assumption that students benefited from doing the first annotated-bibliography exercise. As a group, they got better results (approximately half a grade better) than those who did not do this exercise. In 2018, the annotated-bibliography exercises were a compulsory learning activity; thus, all participating students did it. The average grade among these students was 3.26, thus slightly better than those who did not do the annotated-bibliography exercises in 2017 but not as good as those who did these exercises. Although the intervention of making the annotated-bibliography exercises compulsory provided some improvements, other incentives are needed to significantly increase the level of writing skills for geography students at the bachelor level. Results from the survey and the qualitative interviews provide some ideas for how, and we will return to this issue in the discussion section.
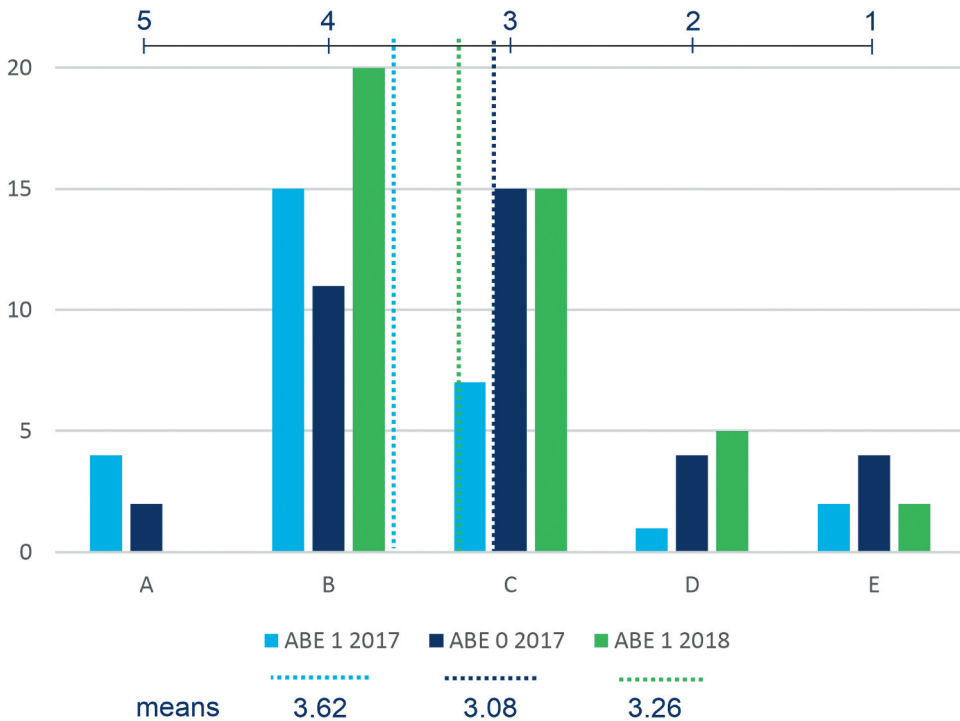
Figure 1. Distribution of grades for students who have done the first annotated-bibliography exercise in 2017 (ABE 1 2017) and those who have not (ABE 0 2017). In 2018, this was compulsory; therefore, all the students did this exercise.

Reading behaviour

Regarding reading behaviour (Table 1), the responses from the survey on statements 1 and 2 seem to indicate that we did provide extrinsic incentives for students to read the assigned material and to search for literature relevant for their own term paper. For statement 1, however, about 36% of responses were neutral to or disagreeing that the term-paper assignment was an important incentive to read the assigned material. From the interviews, we learned that this might be due to two reasons. First, incentives to spend time on the assigned reading were mostly stimulated by the nine tests or mini-exams (not reported here). Second, students found it more important to search and read relevant and recent research literature than to read the assigned material while preparing their term paper, which may explain why the percentage of neutral or disagreeing responses is much lower for statement 2.

For statements 3 and 4, students responded by comparing the reading for the term paper and the three reviews, respectively, with the reading for a traditional end-of-semester exam. Whereas the responses to statement 3 reached a majority on the positive side, no such majority was found for statement 4. This may indicate that students worked more on their own term paper than on the three peer reviews. We found support for this in the responses to an open question in the survey where students were asked to provide comments to questions 1 to 4 in Table 1.

I did not read a lot of literature for the three other topics, just a few articles.

> I learned a lot about my own subject and less about the others, because it takes a lot of time to read that much about many topics.

More variation in responses for statement 4 is also evident from the responses to the open questions where we received varied feedback.

(1) Positive feedback: This is my favourite part of the module, doing term-paper instead of a final exam. I feel that I could learn more because it is more in depth, and we have the luxury of time to slowly read and understand.
(2) Indifferent feedback: You have to read and understand the literature, no matter if it is for an exam or a term paper, so that does not make any difference.
(3) Negative feedback: The curriculum itself is very interesting, but in the way the course is now, you almost do not need to read or learn anything about this curriculum. An exam would have been much, much better than a term paper. Houghton's book Global Warming is great, but with today's setup of the course, you do not need to open it at all. Please go back to having an exam!

According to Sampson and Cohen (2001), when introducing peer learning strategies, an acceptance that different students will learn different things is needed. At least one student seems to have recognised this in his/her response.

> If you have an exam at the end of the semester, you have to learn everything by heart, because everything could be relevant. This does not mean that we learned less, but that we all learned about different aspects of our topics, because even within one topic, you can focus on different parts.

From the interviews, all informants reported that, when they got extrinsic incentives to read, more reading happened. They also found reading for their term paper to be a different experience than reading for a traditional written end-of-term exam. They enjoyed learning about how to search for research literature and the freedom of choosing themselves which articles were relevant for their term paper. The informants read a lot for their term paper and found this reading motivating, for instance, as expressed by one of them.

> For this subject, the list of assigned reading was rather short, but I needed to read a lot for the term-paper assignment, more than what I am used to from other subjects. The topic for my term paper was very interesting, so I found it exciting to search for relevant research literature I could use and to read these articles.

Also, when comparing reading for an exam and reading for a term-paper preparation, the informants emphasised that reading for their term paper represented a much more targeted sort of reading where they could apply the knowledge they gained. Below are two illustrative statements.

> For a traditional written exam, you do not know what questions you may get. When I read for an exam, I therefore try to read something about everything. Reading for a term paper is different. I wanted to know a lot about one topic, not broadly like for a traditional exam. The reading you do while working on a term paper is thus much more targeted. You look for relevant elements and can much more easily drop some if they are not relevant.

> On subjects where you read articles from the list of assigned reading, mostly you read just to know what the article is about, you take some notes, but you do not work with the text more than what is needed to gain some overview. You do not get the chance of applying this work to anything else than possibly being able to answer an exam question well. You may know what is required for the exam, but you do not engage much in the reading, you just skim through a lot of text. When I read for my term paper, and partly for the term-paper peer reviews, I could apply what I had been reading and what I had learned.

All the informants stated that they did not read very much when they did the reviews and that they emphasised structural and grammatical issues when providing reviews more than critical comments about the content.

Perceptions on the peer-review process

Regarding students' perceptions of the review process (Table 2), there was a clear majority of responses stating that the reviews helped to improve their paper and that they were balanced, insightful, helpful, and clear. The responses to questions 5, 6, 7, 8, and 10 were all with a majority on the positive side, and even more so for questions 5 and 8 with almost 80% of the responses on the positive side. A limitation from a closed question survey, however, is that the responses do not tell us in what way the reviews were helpful. Fortunately, the informants provided us with this information. None of the informants received three very elaborate reviews, but they all got one or two very good ones. A positive experience was a broader perspective provided by the reviewers, which helped them to understand the topic and to address the main points. A negative experience was that, although it was required, not all reviewers provided literature recommendation. However, some did it well.

> One of the reviews was very good, providing some good literature recommendations and some critical feedback, which was useful to assess whether my arguments were good enough.

Common for all the informants was the experience that the reviewers mainly pointed at structural and grammatical issues. One of the informants perceived this rather negatively and summarised the peer-review process as follows.

> I did not find the received peer reviews so useful. They mostly corrected spelling mistakes, told me that some paragraphs were short, that I jumped too quickly to some statements, and that an argument was missing. The feedback I received was mostly polishing of my text, its structure, length of paragraphs, or grammar, and not so much on academic content.

For this informant, the peer-review process was not so helpful because her command of English was very good. For most of the other students, writing in English may have been challenging and they may therefore have found the peer-review process helpful because their writing in English improved. This informant found that to be a positive feature.

> I got several comments from the reviews on how I could simplify the text to make it easier to understand by others. Although a sentence made sense for me, I realised after the reviews that it could be difficult to understand for others.

Scores awarded for authoritativeness were without a majority but more dispersed and with a higher score on the neutral category than the other statements (see statement 9 in Table 2). These results correspond to those from Mulder et al. (2014) and Harland et al.

(2017) and may mirror the students' concern about the competence of their peers. We believe that such scepticism may be supportive for developing skills in critical thinking. If students regard feedback from a professor as authoritative, they may be more likely to accept it uncritically. However, if students receive feedback from their fellow students, they may be more likely to be more sceptical, to read the feedback thoroughly, to counter argue, and to reflect about it. Then, students learn more because they are interacting with the feedback process. Some of the open responses seem to support this.

> The reviews were quite different, from very positive to slightly negative. I did not agree with all comments, but it was helpful and can maybe be improved by adding discussion possibilities.

> Sometimes, I simply disagreed with some opinions in the reviews.

From one of the informants, there was also a response that may support that feedback from peer students required much work.

> For some of the comments, I had to work quite a lot to check whether I was wrong or whether the reviewer was wrong. I had to check some things rather well. Sometimes, I was wrong; sometimes, the reviewer had misunderstood my text. I thought this was a good experience.

The variation in the responses for authoritativeness may also be partly explained by the peer reviewer providing few critical comments about the academic content. One of the informants emphasised the academic content of the reviews as more important than the grammatical improvements.

> From the two reviewers with good knowledge, there was some rather good feedback. The academic level of my text was improved, but the language was still my poor English, except for some spelling mistakes they helped me with.

However, there were also reviewers not knowing the topics so well.

> They only partially had good knowledge of the topic. I do not think they had read much to provide the reviews.

> I received three peer reviews, and the reviewers could have deepened their knowledge of the topic more. I did not get any suggestions for research literature I could use or much other feedback on the 'academic level'.

A problem expressed by several of the responses to the open questions from the survey indicated that, for peer review to be successful, the version of the term paper submitted for review must be of a certain quality. Here are two examples.

> As for my feedback-giving experience, I received very incomplete papers to read so it was extremely difficult to provide good feedback. When the paper is missing the conclusions and major parts of the theory/discussion, it is quite impossible to provide "new insight" when you don't see a proper structure and argument yet! I felt more like a teacher rather than a peer, as there were so many things that needed to be fixed in terms of grammar, set-up, or just remembering to use citations on claims from Houghton's book, for example.

> It was difficult to write an appropriate review when the received term paper was not even nearly finished. Perhaps, in the next years, it should be required that at least 2000–2500

words are already written. This would make writing a review definitely more useful! For now, it appeared that the reviewer had to do part of the author's work.

Similar reflections were also expressed by the informants.

(1) Relating to grammar: It is difficult to provide good academic feedback when a text is full of grammatical errors.
(2) Relating to submission of unfinished term papers: One of the problems was that I had not written much for this first submission. Therefore, the reviewers did not have much to provide feedback on. That cannot have been so easy for them.
(3) As well as an example of well-prepared term papers that challenged the reviewer: For the term-paper reviews, I had to catch up with the other topics and had to read some of the material they had read, particularly if they had used one or a few sources thorough the paper.

Discussion and conclusion

Experiences from Australia show that most students perceived double-blind peer review positively and that the quality of students' term papers distinctively improved after peer review (Mulder et al., 2014). We arranged three kinds of seminars preparing the participating students to write their term papers: (1) research-literature search, (2) the first annotated-bibliography exercise, and (3) use of reference-management software to set up a proper reference list. As shown in Figure 1, the students who did the first annotated-bibliography exercise did improve their performance. There is a possible bias, namely, that the sample of students who did the annotated exercises would have received better grades anyway because they may be among the more devoted students. However, the bias could also go the other way, that the best performing students did not need the annotated-bibliography exercise if they already knew well how to search for articles and to write a critical summary of these. Nevertheless, the term-paper assignment was an important extrinsic incentive for students to read the assigned material and, even more so, recent and relevant research articles not on their reading list. This may also have influenced the improvement of term-paper grades. The peer-review assignment, however, did not similarly function as an extrinsic incentive for students to read.

The setup of a double-blind multiple peer review is complicated and requires adaptive software. It is much easier to let students swap their work and provide comments to each other, but this may result in asymmetrical feedback between a proficient and less proficient student (Hanjani & Li, 2014). This is a key reason to arrange multiple peer review, to secure at least some symmetrical feedback between learners of approximately equal skills. From our experience, all students got some symmetrical feedback. Students who receive responses from several peers are therefore more content than two students working together (Mulder & Pearce, 2007).

A key reason to implement double-blind peer review is to prevent the quality of peer review among students from being corrupted by bias due to friendships. A double-blind peer-review process will prevent relationships between students to influence their reviews (Papinczak et al., 2007). Consequently, a double-blind peer review is recognised as a critical factor "to offer students a safe and supportive learning environment in which

they feel comfortable and confident to provide truthful and constructive reviews" (Li, 2017, p. 646).

Our results presented here add support to literature stating that students find peer review useful (Mulder et al., 2014), but the usefulness experienced by students following our setup was mainly related to form (such as structure, grammar, and how to set up a proper reference list) and less related to content. One of the reasons why peer review has a huge potential is because, with peer review, students interact with quality criteria for academic texts. This interaction may be useful for improving the text they are currently writing (such as a term paper) but also for future improvement of general writing skills because they learn to self-evaluate their writing (MacArthur, 2007). We had prepared a peer-review template and rubric that the students used for their peer-review submissions (see Appendix 1). We designed these with the best intention for the students to critically engage with the subject content of the term paper they were writing and with the ones they were reviewing. However, we were probably too naïve because we did not arrange the seminars in a sufficiently targeted manner to train such skills. Consequently, many students may have interpreted critical engagement with academic content as less important, as expressed by one of our informants.

> This was how I interpreted the peer-review tasks, that we were mostly supposed to give feedback on structure, reference techniques, etc., and then, a possible consequence is that the academic content becomes less important.

A lesson we have learned is that one needs to train the students and provide guidance on how to engage critically with academic content. The second annotated-bibliography exercise was aimed to prepare students to this end. Students searched for relevant articles and reflected on why this was relevant for the three topics. This was presented to students in a lecture as an element they should use for the reviews, and we also talked about how to give and receive feedback. We could have done this much more elaborately and involved students more actively. We may therefore, unintentionally, have emphasised structure, grammar, and reference techniques above academic content, because many of the students seemed to have been focusing most on these aspects when providing feedback. We do not know whether this may also have hampered their writing process, but there are critical approaches that argue that traditional formats for term papers (introduction, theory, methods, results, discussion, conclusion) have limited students' abilities to develop a more personal narrative style and develop themselves as successful writers (see for example, Burlingame, 2019; DeLyser, 2010).

According to Carbaugh and Doubet (2016, p. xxi), "one misconception that often accompanies a move to digital learning is the assumption that, because students are online, they are learning . . . In *any* environment – face-to-face or digital – we cannot simply hope that students are processing and reflecting on content; rather, we must guide them to do so". An important way to improve how students may provide each other with critical and constructive feedback on academic content would be to arrange seminars on how this can be done. Hill and West (2020) ran a process they call feed-forward dialogue where students drafted an essay, which was discussed and evaluated with the subject teacher face-to-face. Arranging a face-to-face meeting where the tutor discusses with a student her/his submitted term-paper draft is an ideal approach but difficult to accomplish when teaching resources are scarce. Hill

and West (2020) did the feed-forward dialogue on a course run by one tutor and with 30–45 students but recognised that it was resource-intensive. For a course with about twice as many students, the workload of having face-to-face meetings could be too much, but not if the process is arranged in a way that allows students do this with each other. Face-to-face seminars could be arranged as part of the second annotated-bibliography exercise. In addition to sharing and discussing research literature, students could apply the results from their annotated-bibliography exercise together with the term-paper template and rubric on some sample term paper (for instance, among those submitted by last year's students). The aim of such seminars should be to motivate students while training them in giving peer review and to increase the quality, regarding academic content, of the students' submitted term paper.

Notes

1. One recent and promising exception is Hill and West (2020) dialogic feed-forward assessment.
2. We specified the test as a one-tailed distribution for two samples (non-paired) with unequal variance and used the T.TEST function in MS Excel, which yield $p = 0.02$.

Acknowledgments

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

ORCID

Jan Ketil Rød (iD) http://orcid.org/0000-0003-2935-6206

References

Badcock, P. B. T., Pattison, P. E., & Harris, K.-L. (2010). Developing generic skills through university study: A study of arts, science and engineering in Australia. *Higher Education*, *60* (4), 441–458. https://doi.org/10.1007/s10734-010-9308-8

Balfour, S. P. (2013). Assessing Writing in MOOCs: automated essay scoring and calibrated peer review™. *Research & Practice in Assessment*, 8(1), 40–48. https://files.eric.ed.gov/fulltext/EJ1062843.pdf

Bean, J. C. (2011). *Engaging Ideas* (2nd ed.). Jossey Bass Ltd.

Boud, D. (1995). What is learner self assessment? In D. Boud (Ed.), *Enhancing learning through self assessment* (pp. 11–23). RoutledgeFalmer.

Boud, D., Cohen, R., & Sampson, J. (1999). Peer Learning and Assessment. *Assessment & Evaluation in Higher Education*, 24(4), 413–426. https://doi.org/10.1080/0260293990240405

Burchfield, C. M., & Sappington, J. (2000). Compliance with required reading assignments. *Teaching of Psychology*, 27(1), 58–60. https://psycnet.apa.org/record/2000-07173-017

Burlingame, K. (2019). Where are the storytellers? A quest to (re)enchant geography through writing as method. *Journal of Geography in Higher Education*, 43(1), 56–70. https://doi.org/10.1080/03098265.2018.1554630

Carbaugh, E. M., & Doubet, K. J. (2016). *The differentiated flipped classroom. A practical guide to digital learning*. Thousand Oaks: Corwin, SAGE.

Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98(4), 891–901. https://doi.org/10.1037/0022-0663.98.4.891

Clump, M. A., Bauer, H., & Bradley, C. (2004). The extent to which psychology students read textbooks: A multiple class analysis of reading across the psychology curriculum. *Journal of Instructional Psychology*, 31(3), 227–232. http://web.a.ebscohost.com/ehost/detail/detail?vid=1&sid=a362f49e-f15b-4d64-9025-f1a5e3aeca88%40sessionmgr4007&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#AN=14701522&db=a9h

Cohen, A. J., & Spencer, J. (1993). Using writing across the curriculum in economics: Is taking the plunge worth it? *The Journal of Economic Education*, 24(3), 219–230. https://doi.org/10.1080/00220485.1993.10844794

DeLyser, D. (2010). Writing qualitative geography. In D. DeLyser, S. Herbert, S. Aitken, M. Crang, & L. McDowell (Eds.), *The SAGE handbook of qualitative geography* (1st ed., pp. 341–358). SAGE.

Dysthe, O., & Engelsen, K. S. (2011). Portfolio practices in higher education in Norway in an international perspective: Macro-, meso- and micro-level influences. *Assessment & Evaluation in Higher Education*, 36(1), 63–79. https://doi.org/10.1080/02602930903197891

Ferguson, T. (2009). The 'write' skills and more: A thesis writing group for doctoral students. *Journal of Geography in Higher Education*, 33(2), 285–297. https://doi.org/10.1080/03098260902734968

Gibbs, G. (1999). Using assessment strategically to change the way students learn. In S. Brown & A. Glasner (Eds.), *Assessment Matters in Higher Education* (pp. 41–53). Open University Press.

Gibbs, G. (2006). Why assessment is changing. In C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 11–22). Routledge.

Gibbs, G., & Simpson, C. (2005). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, 2(1), 3–31.

Guthrie, J. T., & Alao, S. (1997). Designing contexts to increase motivations for reading. *Educational Psychologist*, 32(2), 95–105. https://doi.org/10.1207/s15326985ep3202_4

Gynnild, V. (2003). *Når eksamen endrer karakter. Evaluering for læring i høyere utdanning*. Cappelen Akademisk Forlag.

Hanjani, A. M., & Li, L. (2014). Exploring L2 writers' collaborative revision interactions and their writing performance. *System*, 44, 101–114. https://doi.org/10.1016/j.system.2014.03.004

Harland, T., Wald, N., & Randhawa, H. (2017). Student peer review: Enhancing formative feedback with a rebuttal. *Assessment & Evaluation in Higher Education*, 42(5), 801–811. https://doi.org/10.1080/02602938.2016.1194368

Hill, J., & West, H. (2020). Improving the student learning experience through dialogic feed-forward assessment. *Assessment & Evaluation in Higher Education*, 45(1), 82–97. https://doi.org/10.1080/02602938.2019.1608908

Hoeft, M. E. (2012). Why university students don't read: What professors can do to increase compliance. *International Journal for the Scholarship of Teaching and Learning*, *6*(2), 1–19. https://doi.org/10.20429/ijsotl.2012.060212

Hounsell, D., McCune, V., Hounsell, J., & Litjens, J. (2008). The Quality of Guidance and Feedback to Students. *Higher Education Research & Development*, *27*(1), 55–67. https://doi.org/10.1080/07294360701658765

Lei, S. A., Bartlett, K. A., Gorney, S. E., & Herschbach, T. R. (2010). Resistance to reading compliance among college students: Instructors' perspectives. *College Student Journal*, *44*(2), 219–229. http://web.b.ebscohost.com/ehost/detail/detail?vid=0&sid=ca074327-c105-4869-ae94-150387d40569%40pdc-v-sessmgr02&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#AN=51362157&db=a9h

Li, L. (2017). The role of anonymity in peer assessment. *Assessment & Evaluation in Higher Education*, *42*(4), 645–656. https://doi.org/10.1080/02602938.2016.1174766

Lillis, T., & Turner, J. (2001). Student writing in higher education: Contemporary confusion, traditional concerns. *Teaching in Higher Education*, *6*(1), 57–68. https://doi.org/10.1080/13562510020029608

MacArthur, C. A. (2007). Best practice in teaching evaluation and revision. In S. Graham, C. MacArthur, & J. Fitzgerald (Eds.), *Best practice in writing instruction* (pp. 141–162). Guilford.

Mui, A. B., Nelson, S., Huang, B., He, Y., & Wilson, K. (2015). Development of a web-enabled learning platform for geospatial laboratories: Improving the undergraduate learning experience. *Journal of Geography in Higher Education*, *39*(3), 356–368. https://doi.org/10.1080/03098265.2015.1039503

Mulder, R., Baik, C., Naylor, R., & Pearce, J. (2014). How does student peer review influence perceptions, engagement and academic outcomes? A case study. *Assessment & Evaluation in Higher Education*, *39*(6), 657–677. https://doi.org/10.1080/02602938.2013.860421

Mulder, R., & Pearce, J. M. (2007). PRAZE: Innovating teaching through online peer review. ICT: Providing choices for learners and learning. *Proceedings ascilite Singapore 2007*, Singapore. http://www.ascilite.org.au/conferences/singapore07/procs/mulder.pdf.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218. https://doi.org/10.1080/03075070600572090

O'Donovan, B., Price, M., & Rust, C. (2001). The student experience of criterion-referenced assessment (through the introduction of a common criteria assessment grid). *Innovations in Education and Teaching International*, *38*(1), 74–85. https://doi.org/10.1080/147032901300002873

Olin Library Reference. (2018). *How to Prepare an Annotated Bibliography: The Annotated Bibliography*. Cornell University Library. http://guides.library.cornell.edu/c.php?g=32342&p=203789.

Papinczak, T., Young, L., & Groves, M. (2007). Peer assessment in problem-based learning: A qualitative study. *Advances in Health Sciences Education*, *12*(2), 169–186. https://doi.org/10.1007/s10459-005-5046-6

Purchase, H., & Hamer, J. (2018). Peer-review in practice: Eight years of Aropä. *Assessment & Evaluation in Higher Education*, *43*(7), 1146–1165. https://doi.org/10.1080/02602938.2018.1435776

Ryan, T. E. (2006). Motivating novice students to read their textbooks. *Journal of Instructional Psychology*, *33*(2), 136–140. http://web.a.ebscohost.com/ehost/detail/detail?vid=0&sid=55bf0dac-0006-4e30-a147-4747dcc49cd6%40sdc-v-sessmgr03&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#AN=106160368&db=a9h

Sampson, J. & Cohen, R. (2001). Designing peer learning. In BoudD., CohenR., & Sampson J. (Eds.), *Strategies for peer learning: some examples* (pp 35–49). Kogan Page.

Sitthiworachart, J., & Joy, M. (2008). Computer support of effective peer assessment in an undergraduate programming class. *Journal of Computer Assisted Learning*, *24*(3), 217–231. https://doi.org/10.1111/j.1365-2729.2007.00255.x

Smith, S. L. (1982). Learning strategies of mature college learners. *Journal of Reading*, *26*(1), 5–12. https://www.jstor.org/stable/40029209

Søndergaard, H., & Mulder, R. A. (2012). Collaborative learning through formative peer review: Pedagogy, programs and potential. *Computer Science Education*, *22*(4), 343–367. https://doi.org/10.1080/08993408.2012.728041

Taffs, K. H., & Holt, J. I. (2013). Investigating student use and value of e-learning resources to develop academic writing within the discipline of environmental science. *Journal of Geography in Higher Education*, *37*(4), 500–514. https://doi.org/10.1080/03098265.2013.801012

Tsai, -C.-C., & Liang, J.-C. (2009). The development of science activities via on-line peer assessment: The role of scientific epistemological views. [journal article]. *Instructional Science*, *37*(3), 293–310. https://doi.org/10.1007/s11251-007-9047-0

Wind, D. K., Jørgensen, R. M., & Hansen, S. L. (2018, 5-6 July). *Peer Feedback with Peergrade. ICEL 2018-13th International Conference on e-Learning*, The Cape Peninsula University of Technology, Cape Town, South Africa.

Wingate, U., & Tribble, C. (2012). The best of both worlds? Towards an English for Academic Purposes/Academic Literacies writing pedagogy. *Studies in Higher Education*, *37*(4), 481–495. https://doi.org/10.1080/03075079.2010.525630

Appendix 1: Peer review template

Review of term paper # _____ (use the number used as file name)

Topic _____ (fill in A, B, C, or D)

Text written in red font are elements you should use when providing feedback. Replace the red text with your review text but keep the headings (marked with bold). Remember that your review should be constructive and polite. Be specific in your review, use page and line numbers to make it easy for the term-paper author to understand what part of her/his text you are commenting.

Comments on introduction

- How well does the term paper introduce topic A/B/C/D and its background?
- How well does the term paper identify the main points made in the given texts?

(The given texts are the ones written by Mörner, Atkins, Cheetham, or Lomborg.)

- Comment on how and where the term-paper author presents the main points.
- Are some major points missing?
- Are all points clearly stated and well-articulated? If not, where and how could they be improved?

Comments on analysis / discussion

- How well does the term paper critically discuss the main points made in the given text?
- How do the main points relate to research findings? Are there contested issues or is there a consensus among the research community about the main points made in the given text?
- Has the term paper author included relevant research literature for this discussion / analysis?
- Has he/she included literature from the reading list for the course as well as other research literature? If yes, praise them for this in your review text. If not, suggest where and how this can be done; for instance, by bringing in literature you have found from the second annotated exercise. Has he/she included research articles that you have found from your second annotated exercise? If not, and if you mean this/these article(s) are relevant, suggest the term-paper author to include these, why, where, and how.
- How well is the selected research literature integrated in the text to support the arguments / statements used by the term-paper author?

Comments on conclusion

- How well is the conclusion synthesising the main points?

Writing quality and organization

- Is the paper well structured (is the text where it should be or should part of the text be moved)?
- Is the text easy to follow?
- Is the reference list complete (all references in the text are listed in the reference list)?
- Are all references made according to the APA style?

Review of term paper # _____ (use the number used as file name), Topic _____ (fill in A, B, C, or D)

The rubric below is a suggestion for how one could assess a term paper. Please add elements if you think something is missing. If so, please mark your edits (for instance by using red font). Give a score for the term paper by filling in numbers in the yellow markings.

Aspect	Sophisticated	Competent	Needs Work
1. Introduction (_/5)	The term-paper author introduces the text for topic A/B/C/D properly (i.e., author and source). All of the main points of the term paper are clearly stated. 4–5 pts	The term paper author introduces the text for topic A/B/C/D properly. The main points of the term paper are clear in parts or only partially described. 2–3 pts	Introduction is incomplete, poorly organised, and/or vague. 0–1 pts
2. Research and analysis (_/10)	Selected research literature is highly relevant to the arguments and comes from highly reputable sources. Arguments are presented accurately and thoroughly, and the referred literature is very well integrated with the term-paper author's arguments. 8–10 pts	Selected literature is relevant to the arguments but too little from peer-reviewed research literature. Arguments are mostly presented well, but there are some unclear parts and/or some minor errors in the presentation. The referred literature is, for the most part, well integrated with the term-paper author's arguments but has some loose ends / minor errors. 5–7 pts	All or most of the literature presented is not relevant to the argument and/or is hardly from reviewed research literature. Arguments are not clearly articulated and/or have incorrect or incomplete components. Major errors in the logic are present. The referred literature is only vaguely, incompletely, or inaccurately integrated with the term-paper author's arguments.0–4 pts
3. Conclusions (_/5)	The conclusion is a clear and concise synthesis of the main points. The main points are very well restated and summarised. The underlying logic is explicit. No new arguments or references are introduced. 4–5 pts	The conclusion is a good synthesis of the main points but could have been more clear and/or concise. The main points are well restated and summarised but are somewhat repetitive of the findings and/or introduce new arguments/references. Minor errors in logic are present. 2–3 pts	The conclusion is a poor synthesis of the main points. The main points are poorly restated and summarised or are just a repetition of the findings without explanation, and/or several new arguments or references are introduced. Underlying logic has major flaws. 0–1 pts

(*Continued*)

(Continued).

| 4. Writing quality and organisation (_/5) | The paper is coherently organised, and the logic is easy to follow. There are only minor spelling or grammatical errors, and terminology is clearly defined. Writing is clear, concise, and persuasive. References are complete and in proper format in both the text and the reference list. 4–5 pts | The paper is generally well organised, and most of the argument is easy to follow. There are some spelling or grammatical errors, or terms are not clearly defined. Writing is mostly clear but may lack conciseness. References are complete and in proper format in both the text and the reference list. 2–3 pts | The paper is poorly organised and difficult to read. It does not flow logically from one part to another. There are several spelling and/or grammatical errors, and technical terms may not be defined or are poorly defined. Writing lacks clarity and conciseness. References are incomplete (either in text and/or in reference list) and not properly formatted. 0–1 pts |

Total points (_/25)