Is Peer Feedback Helpful When Learning Literature Review Writing? A Study of Feedback Features and Quantity

Evelyn Eika

Correspondence: Evelyn Eika, Faculty of Technology, Art and Design, Oslo Metropolitan University, 0130 Oslo, Norway.

Received: March 1, 2021 Accepted: March 24, 2021 Online Published: March 27, 2021

doi:10.5430/elr.v10n1p10 URL: https://doi.org/10.5430/elr.v10n1p10

Abstract

The literature review is an important part of an academic text. It is common for students to learn literature review writing through reading and practice. This study explored peer feedback as an assisting interactive social tool that aims to provide formative assessment of literature reviews written by master students. The student reviewers' feedback characteristics were identified and coded, and their relationships with writing performance were examined. Six assessment criteria and writing categories relevant to the literature reviews genre were employed for the writing process as well as a peer feedback process. The feedback patterns were analysed according to four dimensions: describing status or problems versus prescribing directions, abundant input versus uncritical/empty comments, high versus low level, and constructive versus negative. No correlations were found between review patterns, students' performance mark, and quantity of feedback received. Significant correlations were observed between specific review patterns and separate category scores. The dimensions constructive vs. negative and high vs. low level correlated with most category scores. The findings show that the students were able to provide useful and high-level comments to assist their peers' writing. Overall, it was found that peer feedback quality and quantity do not define the performance mark, but benefit individual aspects of literature reviews writing.

Keywords: higher education, academic English, research students, literature reviews genre, peer feedback pattern, writing rubric, revision

1. Introduction

The literature review, or related works section, is a key component of an academic text. The process of writing a literature review can help the author discover related work, understand the research discourse, and reflect over his or her research in a larger context, thereby justifying the research efforts. To learn the genre of literature review writing is therefore paramount for research students.

English has become the universal academic language (Phillipson, 2017). Investigations into English research articles writing have focused on different separate sections, including abstracts (Xie, 2020; Tanko, 2017; El-Dakhs, 2018; Khedri, Heng, & Ebrahimi, 2013), introduction (Kawase, 2015), results (Basturkmen, 2009), discussion (Hopkins & Dudley-Evans, 1988), conclusion (Bunton, 2005), and global structures of doctorate dissertations (Anderson, Alexander, & Saunders, 2020). Very little research has dealt with literature reviews writing as a genre. Literature reviews refer to the type of article functioning as a tutorial, identifying the problem, describing relevant scholarship in terms of concepts and methodologies, clarifying inconsistencies and contradictions, and suggesting future research directions (The American Psychological Association, 2009, p. 10). This study explores how beginner-level research students develop their literature reviews.

To engage students in learning through writing, peer feedback has been considered beneficial in the context of higher education (Hanrahan & Isaacs, 2001). Providing feedback to peers and utilising feedback from peers are also important skills that would benefit students in their academic performances (e.g., by fostering development of subject-specific skills, van Zundert, Sluijsmans and van Merriënboer, 2010) and for their future professional career (Huisman, Saab, van Driel, & van den Broek, 2017). With increased massification of higher education, student diversity and class sizes continue to increase (e.g., Anderson, Alexander, & Saunders, 2020); employing peer assessment would help support individuals' learning needs and enrich learning experience (Vickerman, 2009). A recent meta-analysis revealed that students who engaged in peer feedback improved their writing more than not engaging in any feedback (peer, self, or teacher), but the study concluded that its generalisability was limited due to

limited quantitative studies (Huisman et al., 2019). However, what seemed to be certain is that the greater number of peers involved, the more quality peer feedback would be obtained, hence helping to improve writing quality Such positive impact has largely been attributed to newness or differences that peers may potentially bring to writers including different perspectives, approaches, and writing styles (McConlogue, 2015). As feedback providers, students may be forced to actively consider assignment criteria which may in turn help improve their own writing performance (Flower et al., 1986; Patchan & Schunn, 2015). Employing peer review could also serve as instructional purposes as students would learn from their peers and become aware of how their own work may be perceived (Stellmack et al., 2012).

Concerning peer feedback quality, summative views considered accuracy and consistency across peers and teacher feedback. For instance, Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh (2010) examined feedback effectiveness on first-year business undergraduates' letter writing that contained structure and content flaws. Criteria employed included calculating instances of errors uncovered as opposed to total number of errors, error correction and justification, as well as an overall score illustrating peer comments in terms of correctness, specificness, and comprehensiveness. Another perspective focused more on feedback composition characteristics. For example, Sluijsmans et al. (2002) employed seven characteristics based on expert opinions to examine teacher students' peer feedback: the number of constructive criteria used (such as clear, direct, accurate, comprehensible, and achievable), positive comments, negative comments, constructive suggestions, posing questions, and no use of imprecise words (e.g., good, nice, and fine). However, they did not investigate the relationship between these feedback characteristics and their effectiveness. Prins et al. (2006) further developed the characteristics used by Sluijsmans et al. (2002) into a Feedback Quality Index for medical consultation contexts. This index employed weighted category differences instead of occurrence counts: content-based remarks, remarks explanations, suggestions for improvement, balance of positive and negative comments, reflective questions, clear descriptions (instead of key words), structure, examples, and style (using first person instead of judging). Effectiveness of specific feedback characteristics onto performance was again not explored. Kim (2005) explored the relationship between providers' feedback characteristics and receivers' performance. Peers gave feedback based on assignment requirements on concept maps by checking yes or no and leaving comments as well as choosing one of three pre-written statements regarding structure, topic coverage, amount of support material, and creativity, finalising each with a reason for their choice. However, no performance improvement was found, and most peer assessors provided no suggestions for revision. The study deemed that students' inability to internalise feedback may be the cause. Gielen, Peeters, et al., (2010) employed seven feedback characteristics to examine their impact on essay assignments written by secondary school children appropriateness (based on assessment criteria), specific explanation, justification (reasons for the comments), suggestion for improvement, clear formulation, both positive and negative comments, and thought-provoking questions. The last two characteristics were later omitted because of only one occurrence and assessment difficulty. Only justification was found to have a significant effect on writing improvement, but the impact was reduced for learners who had high pre-test performance. Appropriateness and clear formulation exhibited the highest mean score while specificity and justification the lowest, suggesting that grade 7 learners were able to provide appropriate and clear comments but found it hard to give specific explanation or justification. However, providing justification was far more effective than peers' comment accuracy. Using online peer assessment to examine its impact on macroeconomics essay performance, Snowball & Mostert (2013) found that students made more useful comments and suggestions on presentation and referencing, followed by structure and organization, then language and requirements (such as number of references cited and length), and lastly on content-related issues. Participating in peer feedback, however, did not significantly determine the final essay mark, but students' subject-specific ability and English language proficiency did. A recent study by Huisman, Saab, van Driel, & van den Broek (2017) investigated first-year undergraduate students' ability match in feedback characteristics and essay performance. Same-ability peers and different-ability peers were asked to provide feedback on content, structure, and style. Comments types were classified as analysis (concerning text meaning), evaluation (quality statements), explanation supporting evaluations or suggestions, and revision (explicit suggestions for text improvement or implicit ones for improvement direction). They found that individual ability did not define feedback quality, and peer feedback quality was not related to writing performance. Still, the participants benefited similarly for all the included aspects of writing. Comparatively high-ability peers provided more content-related feedback. Analytical comments were rare; suggestions for improvement were more frequent and mostly concerned writing style and less on content and structure which were generally more evaluative.

Few studies examined empirical association between peer feedback characteristics and writing performance, and only with mixed results on performance improvement. Only justification comment showed a significant effect on

writing gains (Gielen, Peeters, et al., 2010). However, an observed phenomenon appeared to be that peer reviewers tended to give more comments concerning writing style than content-related or structural aspects of writing (Snowball & Mostert, 2013; Huisman et al., 2017), but high-ability peers provided more content-related feedback (Huisman et al., 2017). This study explored whether peer feedback features do enhance writing performance for beginner-level master-level students who could be considered new to research and research writing but advanced in terms of learning Also, the writing task involved the literature reviews genre. Literature reviews are present in all disciplines and research fields and would thus help to prepare research students potential future publishing activities Writing ability and reviewing ability may be interrelated as there exist certain overlaps in cognitive processes between writing and providing feedback on writing, such as defining tasks, uncovering and diagnosing problems, and revising strategies (Patchan and Schunn, 2015). Hence, it could be hypothesized that high-ability writers are also high-ability reviewers (Huisman et al., 2017). A goal of this study was therefore to explore the relationships between characteristics and effectiveness of peer feedback for research students' literature survey writing. Peer feedback effectiveness was investigated through writing performance in both the overall mark and different writing aspects. Specifically, the study aimed to (a) explore peer assessors' feedback pattern in assessing their peers' manuscript in all respects, (b) determine what writing aspects benefit the most in terms of peer reviewers' feedback that aims to assist student writers revise their work, (c) explore how feedback in terms of quality and quantity affect learning outcomes measured in final grades, and (d) probe the relationship between the academic performance of the peers and the feedback they provide (i.e., do research students with higher academic performance provide more useful feedback?). The following research questions were therefore formulated:

RQ1: What review patterns do different writing categories exhibit?

RQ2: Are review patterns related to writing category scores?

RQ3: Are student writers' final grades related to review patterns received and/or the quantity of reviews received?

RQ4: Are student reviewers' grades related to the review patterns they provide?

2. Method

2.1 Rubrics to Assess Writing

This work focuses on peer feedback content and patterns for the genre of literature reviews APA considers this genre a tutorial in that it defines the problem, describes relevant studies to inform readers of the status of research, discusses the relationships among related studies, and suggests future research directions (APA, 2009, p. 10). To help guide new research students' literature reviews writing and also assist assessors' assessing their manuscript, a six-rubric scheme was proposed: content, structure, text quality, novelty (original perspectives), timeliness (topic appropriateness), and references (Anonymous author, in review). Content addresses subject understanding, intended objectives, theory developed, and conclusions gathered. Novelty concerns new perspectives derived from the current literature. Structure deals with material organization, including section signposting and display of visual information. Text quality concerns expression clarity and consistency. References reflect a writer's quality of sources obtained, indicating newness of material gathered, which is an important element of research and hence a vital part of the literature survey genre. Timeliness is related to both current research and future research as it concerns the topic area that a writer chooses being up-to-date and relevant, an important trait of technology-related research.

2.2 Peer Feedback Dimensions

To analyse the student reviewers' comments, the following key dimensions were identified after reviewing the literature (see Table 1): describing states versus prescribing directions (D vs. P), abundant feedback versus uncritical/empty comments (A vs. E), high level versus low level (H vs. L), constructive versus negative (C vs. N). Thus, there are 16 possible patterns (see Table 2) based on the four dimensions. These dimensions are addressed in the following:

Table 1. Review dimensions

Description	Abbreviation	Examples	
Describe states/prescribe directions	D/P	D: Gives good understanding overall of the use of C for disaster recovery	P: Some more explanation needed in the final discussion section on comparisons with A and B versus C would make it more clear to the reader
Abundant (plentiful, adequate)/uncritical (empty, ignorant)	A/E	A: Explain a bit more about how to maintain the plan and how to implement the plan, point out some more of the risks and elaborate some of these more	E: It is a good idea to use pdf first of all, and it will be better if you can use some pictures
High-level/low-level	H/L	H: You have covered the different aspects about wireless security, strengths and weaknesses with the different protocol	L: By the way this word "witch" still exists in some parts
Constructive/negative	C/N	C: Supported by good references	N: Overall very poor

Table 2. Review pattern taxonomy

		Describe		Prescribe	
		Abundant	Empty	Abundant	Empty
High-level	Constructive	DAHC	DEHC	PAHC	PEHC
	Negative	DAHN	DEHN	PAHN	PEHN
Low-level	Constructive	DALC	DELC	PALC	PELC
	Negative	DALN	DELN	PALN	PELN

Prescribe or Describe (P/D): Suggestions for improvement is a key goal of the peer feedback process. However, the terminology used varies: e.g., solicit clarifications, identify problems, explain problems, and suggest solutions (Stanley, 1992; Tang & Thitecott, 1999); error correction and justification (Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh, 2010); analysis or questioning text meaning (Huisman et al. 2017), justification or explanation (Kim, 2005; Sluijsmans et al., 2002; Prins, Sluijsmans, & Kirschner, 2006; Gielen, Peeters, et al., 2010; Huisman et al., 2017); improvement suggestions (Kim, 2005; Sluijsmans et al., 2002; Prins, Sluijsmans, & Kirschner, 2006; Gielen, Peeters, et al., 2010); evaluation or quality statements (Huisman et al. 2017); remarks explanations (Prins et al., 2006); revision (Huisman et al. 2017). As this study addresses beginner-level research students' writing on literature survey, the pair term is proposed to help capture reviewers' comment tendency: prescribing directions for improvements versus describing current status without suggestions.

Abundant or Empty (A/E): Several studies on peer review have investigated the depth or appropriateness of student reviewers' comments. The approaches used in these studies vary: for example, number of errors uncovered (Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh, 2010), number of criteria used (Sluijsmans et al., 2002), no use of imprecise words (Sluijsmans et al., 2002), appropriateness (Gielen, Peeters, et al., 2010), and clear formulation (Gielen, Peeters, et al., 2010), and usefulness (Snowball & Mostert, 2013). The abundant-empty terms are used in this study to capture the depth of student feedback.

Constructive or Negative (C/N): Previous studies have also investigated the student reviewers' comments being positive or negative. For instance, the studies on positive comments, negative comments, and constructive comments (Sluijsmans et al., 2002); balance of positive and negative comments (Prins et al., 2006); positive comments and negative comments (Gielen, Peeters, et al., 2010). The pair term constructive versus negative is used herein to avoid confusing positive with prescriptive (P).

High or Low (H/L): This dimension was included based on the suggestion (Vickerman, 2009; Cho & MacArthur, 2010; Huisman et al., 2017) that high versus low level remarks seemed to reflect student reviewers' experience or proficiency.

2.3 Participants

First-year master students (n = 21) studying for a technology-related degree participated in the study. The students had various cultural and language backgrounds. All the students used English during official class hours and all satisfied a minimum English proficiency (TOEFL score ≥ 550, or IELTS score ≥ 6.0). None of the students reported any hearing, speaking, or learning disabilities.

2.4 Procedure

2.4.1 Writing Task

The students were to write a literature review of 3,000 words in English (excluding references) based on a topic of their own choice within the thematic scope of the master programme. The students were instructed and encouraged to consult the APA (6th version, 2009) publication manual. Three drafts were to be submitted with two opportunities of written peer feedback; each feedback was followed by authors' revisions prior to final submission to improve quality. Authors were also to construct response memos to respond to all feedback received. All were completed with pre-determined due dates.

2.4.2 Review Task and Response Memo

Prior to each review session, the peer feedback process was instructed and practised during class sessions. The purpose and benefits of peer feedback were firstly explained and emphasized to all students. The peer review form was then distributed to all students and instructed item by item. The review form contains all the six writing criteria (see the first section of method), each with instructions and prompting questions to help reviewers reflect and assess writing. A reminder was also stated in the review form that peer reviewers were to be as constructive and concrete as possible, not criticizing for the sake of having to complete the review task. The reviewers and authors were known to each other and their names were to be listed on the top of the review form. For the practice session, the authors brought two copies of their printouts to class and exchanged their manuscripts with two other peers. A triad was then formed where each manuscript would be orally commented upon by two reviewers. One reviewer would read aloud and comment on the work while the author was to listen and take notes without defending remarks; the third peer within the triad would then use the time to read the manuscript silently. As a practice run, the students were instructed to focus on the quality first, not worrying about the amount of material they must read through in class. During the activity, the teacher served as a facilitator and oversaw the process, providing assistance if students had questions while also allowing participants autonomy. If the triad did not manage to finish reading and commenting on all three manuscripts, the students were encouraged to continue and complete the task outside class time. The oral commentary practice was believed to be useful for writers as they could hear from the commenters reading aloud and think aloud and speaking about their writing, which may help writers form metacognitive strategies in self-evaluating and revising their writing via self-monitoring during/while listening (cf. DiPardo & Freedman, 1988; Tang & Thitecott, 1999). Further, by demonstrating students' thinking processes (Bransford et al., 2000), any thought thus visible may be employed as metacognitive formative evaluation for refining and modifying learners' thoughts to correct biases, misconceptions, and inconsistencies. Such verbal commenting and receiving would in effect be useful for both student reviewers and student authors as the task activates their cognition and enlighten them to form new goals and develop knowledge base (Flavell, 1979). This socio-cultural approach is also recommended for assisting learners in developing their thought (and language) through social interactions in the environment given (Vygotsky, 1986). After the practice session, the authors sent their manuscripts to three peer reviewers who were randomly assigned. The reviewers had about one week to complete their review and document their feedback based on the review form. The peer reviewers were then to send to the authors the feedback they provided. The authors were to revise their manuscript based on the review reports they received. In addition, the authors were to construct a response memo addressing the comments received. For each comment, the author was to respond as to whether any action was done as the peer reviewer suggested or done differently than the peer suggested, or not done and why (justification). The process was administered a second time for the second round of peer review. The authors were to revise their work based on the peer feedback received and then resubmit the text as second draft. A second practice session followed after the second draft was completed. The authors were then to send their second version to the three same reviewers. The reviewers were then to complete the second round of review and send their feedback to the authors. The authors were to revise their manuscript based on the second round of review reports and submit as the third and final draft as well as documenting their changes or non-changes with justifications in a second response memo.

2.4.3 Teacher Grading

The literature reviews were rated by two teacher raters using the same six assessment criteria. Each category was rated using a 1-5 Likert scale. As the first version of the literature survey was considered a practice run, this draft was not rated by the teachers. Hence the second draft was rated as the first version, and the final draft was rated as the second version. This last version was also marked for a collective and final grade, using a scale from A to F where A is the top score and F is a fail.

2.5 Analysis

Contingency tables and Spearman correlations were used to analyse the results. Continency tables were chosen as these are the established way of statistically analysing interactions between categorical data. Spearman correlations are considered appropriate for the analysis of ordinal observations. Statistics were computed using the JASP software package version 0.8.6.0.

Table 3. Feedback patterns frequencies (Only occurring patterns are listed.)

Pattern	Content	Novelty	Structure	Text quality	References	Timeliness	Total
PELC	12.7%	1.6%	26.2%	15.1%	24.6%	1.6%	13.6%
DELC	21.4%	11.1%	11.9%	24.6%	16.7%	15.9%	16.9%
DEHC	7.1%	10.3%	0.8%	3.2%		19.8%	6.9%
PAHC	12.7%	7.9%	6.3%	3.2%		1.6%	5.3%
PEHC	7.1%	9.5%	4.0%	1.6%			3.7%
PALC					5.6%	0.8%	1.1%
DEHN	0.8%						0.1%
PELN				0.8%			0.1%
No comment	38.1%	59.5%	50.8%	51.6%	53.2%	60.3%	52.2%

3. Results

3.1 RQ1: What Review Patterns Do Writing Categories Exhibit?

Table 3 lists the frequency of the reviewing patterns. The reviewers employed only eight patterns, namely DELC, PELC, DEHC, PAHC, PEHC, PALC, DEHN, and PELN. DELC was used most commonly (total 16.9%, highest in text quality 24.6% and lowest novelty 11.1%): *Descriptive*, *Empty*, *Low* level, and *Constructive* comments.

PELC (total 13.6%, highest in structure 26.2% and lowest novelty/timeliness 1.6%) was the second most used pattern denoting *Prescriptive*, *Empty*, *Low* level, and *Constructive* remarks. The difference between PELC and DELC was only prescriptive vs. descriptive. The reviewers were able to give prescriptive input, that is, to offer advice as opposed to simply describe what the author has done in praise form or describe a problem without advice for improvement. It is also noted that prescriptive elements often follow descriptive elements (e.g., "good content, good understanding, explain a bit more about how to maintain the plan and how to implement the plan").

DEHC (total 6.9%, highest in timeliness 19.8% and lowest structure 0.8%) was the third most used pattern. The reviewers offered *Descriptive*, *Empty*, *High* level, and *Constructive* comments. This pattern showed that the student reviewers were indeed able to provide high level or abstract insight that may help strengthen the manuscript along the conceptual dimension, in clarity, or depth of thoughts. One example is this:

> The quality of the text is high, and holds a high consistency and readability through the paper. The references list at the end is well written and follows the APA standard. It is very good that you have not only reference to the article but also where in the article you found what you have used. Security questions in Computer science can quickly change through time, so since article number 2 is from 2001 be sure to check if it somebody or if there is a newer article or some changes since they wrote it.

PAHC was the fourth most (total 16.9%, highest in text quality 24.6% and lowest novelty 11.1%) commonly used pattern. The reviewers provided *Prescriptive*, *Abundant*, *High* level, and *Constructive* remarks. This pattern demonstrated that several students were able to offer plentiful high level and constructive suggestions. Below is one example:

> This is a well written paper. You have covered the different aspects about wireless security, strengths and weaknesses with the different protocols. MAC filtering: mac addresses of devices connected to a wireless access point can be obtained with different software. MAC spoofing is also easy to do, so MAC filtering provides a very limited security. Passwords: you could mention passwords as part of attack on WLAN security. What makes a good password, and what is an insecure one. Just a suggestion. Overall a good paper!

PEHC was the fifth most used pattern (total 16.9%, highest in text quality 24.6% and lowest novelty 11.1%). The reviewers offered *Prescriptive*, *Empty*, *High* level, and *Constructive* remarks. One extract is this: "Please identify your topic on the first cover page. It says literature survey but does not say anything about the topic of discussion. A table chart that shows what the different network analysis tools can do would be helpful."

PALC was the sixth most used pattern (total 1.1%, highest in references 5.6% and lowest timeliness 0.8%). *Prescriptive*, *Abundant*, *Low* level, and *Constructive* comments mostly occurred in the references category. The reviewers were able to give considerable advice in references related issues, reflecting individual knowledge of references that may help strengthen research sources. One example is as below:

> The reference list is well written. And it is very long, showing you have read a lot. Some of the books look kind of old though, like [4] from 2000, but a lot of them are new, form 2013, etc.

DEHN (total 0.1%, content 0.8%) and PELN (total 0.1%, text quality 0.8%) were the least used. DEHN only occurred in the content category, suggesting that *Descriptive*, *Empty*, *High* level, and *Negative* pattern was comparatively less used. With this pattern, reviewers provided descriptive and high-level remarks involving global aspects but considered superficial. One example is listed: "This is more like a market analysis report, I am totally attracted by it since I am always hoping make an app of NFC. But it is not that good as a science paper."

The pattern PELN only occurred in the text quality category. The reviewers gave *Prescriptive*, *Empty*, *Low* level, and *Negative* remarks. As shown, PELC was more frequently used than PELN, differing only in constructive vs. negative. One extract is this: "Overall very poor, need checkup."

The fewest comments occurred for the novelty (59.5%) and timeliness (60.3%) categories, possibly a reflection of lack of interest or reviewers' low priority in the two categories. Unfortunately, more than half of the reviewers (52.2%) provided selective comments, not answering all categories necessarily. The most answered category was content (61.9%) and least answered timeliness (39.7%).

Table 4. Changes in aggregated review patterns from the 1st to the 2nd revision (frequencies)

Pattern	Content	Novelty	Structure	Text quality	References	Timeliness	Sum
DELC	13	6	5	15	3	6	48
DEHN	-1						-1
PELN				-1			-1
PAHC	-4		-2				-6
PELC		-2	7	-11	-1		-7
PEHC	-9	2	-3	-2			-12
No comment	2	-5	-8	-1	-3	-6	-21

3.1.1 Changes in Aggregated Review Patterns From First to Second Version

Table 4 shows the difference in occurrences of patterns from the first to the second revision. For content, PEHC (-9 instances) and PAHC (-4 instances) changed to DELC (+13 in version 2). DELC patterns increased the greatest compared to the other patterns (+48) across categories. This change is positive as *Prescriptive*, *High* level, and *Constructive* comments that mostly occurred in version 1 have changed into more *Descriptive*, *Low* level, and *Constructive* remarks in version 2, suggesting that writers have improved their writing based on the reviewers' suggestions and hence there tended to be more descriptive and lower level comments in version 2.

For structure category, more reviewers provided more comments in version 2 than in version 1 (no-comments exhibited the largest reduction of -8), suggesting more material in version 2 than in version 1 as expected and observed since students had developed more insight and extended the text. Reviewers provided *Prescriptive* and

High level comments in version 1 (PAHC, PEHC), while both *Prescriptive* and *Descriptive* (not just *Descriptive*) but *Low* level (+12) comments in version 2 (PELC, DELC). Version 2 also contained fewer instances of *Abundant* comments (-2). All comments were *Constructive* in both versions.

Concerning text quality, peers provided *Prescriptive Low* comments in version 1 (+11, PELC) but *Descriptive Low* in version 2 (+15, DELC). *High* level comments only occurred in version 1 (PEHC) and only one *Negative* comment occurred in version 1 (PELN). Novelty, references, and timeliness exhibited *Descriptive Low* comments (DELC) in version 2, also with more comments provided in the new revision (no-comments reduced 5, 3, and 6 instances respectively). Novelty additionally contained *Prescriptive High* comments (+2, PEHC) and fewer *Prescriptive Low* comments (-2, PELC) in the new revision. The results showed a more complete writing (i.e., in version 2) triggered more reviewers' novelty comments, despite there being more *Descriptive Low* and less *Prescriptive High* comments.

Table 5. Changes in reviewer feedback from the 1st to the 2nd revision (frequencies)

Feature	Content	Novelty	Structure	Text quality	References	Timeliness
Prescriptive	-13		2	-14		-1
Descriptive	11	5	6	15	3	7
Empty	4		2		-1	1
Abundant	-4		-2		1	-1
High-level	-15	1	-4	-2		1
Low-level	13	4	12	3	3	5
Constructive	-1	5	8	2	3	6
Negative	-1			-1		

3.1.2 Changes in Reviewer Feedback From First to Second Revision

Table 5 shows the difference in occurrence of the features for the six writing categories from the first to the second revision. Content and text quality shifted from *Prescriptive* (-13) to *Descriptive* (+11) across the two rounds of review, indicating more prescriptive input for the first round of reviews while the second round was more descriptive. Content and structure shifted from *High* (-15 and -4) to *Low* (+13 and +12), supporting the view that high-level input is needed in earlier drafts to help writers increase quality and quantity in writing while a low-level comment occurs more frequently for later versions as high-level feedback may have been dealt with during the revision stage.

There were few changes between *Empty* and *Abundant*, ranging from 4 to -4 instances, possibly signalling individual reviewer consistency for the empty vs. abundant dimension independent of writing performance and versions. Observations of the dimension *Constructive* versus *Negative* showed that reviewers tended to be encouraging and hence generally provided more positive (rather than destructive) comments for all categories (negative instances reduced, -1). In fact, there was only one instance of *Negative* (DEHN) in content of version 1. This text was indeed immature and incomplete in all categories; thus, explaining the reviewer's direct and seemingly harsh remarks.

Table 6. Interaction between the rubrics and the review pattern dimensions. Contingency table results for rater A, manuscript version 1

		Content		Novelty		Structure		Text quality		References		Timeliness	
		X^2	p	X^2	p	X^2	p	X^2	p	X^2	p	X^2	p
Descriptive	Content									3.44	.328	2.57	.463
-Prescriptive	Novelty												
	Structure									5.16	.16		
	Text quality									7.51	.057	4.07	.254
	References									3.70	.296	1.48	.687
	Timeliness									2.37	.499	0.74	.865
Abundant	Content									2.42	.491	1.37	.714
-Empty	Novelty												
	Structure									4.38	.224	14.893	.002**
	Text quality									2.00	.572	4.49	.213
	References									1.61	.657	2.49	.477
	Timeliness									1.87	.601	0.90	.826
High-Low	Content									2.02	.569	3.44	.329
	Novelty												
	Structure									3.51	.32	21.64	< .001***
	Text quality									0.67	.881	1.19	.755
	References	25.29	< .001***	44.00	< .001***	20.74	< .001***	11.45	.01**	8.36	.039*	16.61	< .001***
	Timeliness									2.69	.442	5.57	.135
Constructive	Content	32.26	< .001***	44.90	< .001***	24.74	< .001***	17.37	< .001***	10.63	.014*	29.58	< .001***
-Negative	Novelty	28.79	< .001***	50.93	< .001***	22.00	< .001***	11.71	.008**	12.00	.007**	21.43	< .001***
	Structure	23.71	< .001***	49.14	< .001***	24.771	< .001***	12.20	.007**	13.34	.004**	21.57	< .001***
	Text quality	24.65	< .001***	47.23	< .001***	20.74	< .001***	10.68	.014*	8.36	.039*	16.61	< .001***
	References	25.29	< .001***	44.00	< .001***	20.74	< .001***	11.45	.01**	8.36	.039*	16.61	< .001***
	Timeliness	20.93	< .001***	42.00	< .001***	20.86	< .001***	10.00	.019*	7.71	.052	15.14	.002**

Table 7. Interaction between the rubrics and the review pattern dimensions. Contingency table results for rater B, manuscript version 1

		Content		Novelty		Structure		Text quality		References		Timeliness	
		X^2	p	X^2	p	X^2	p	X^2	p	X^2	p	X^2	p
Descriptive -Prescriptive	Content	6.31	.043*	1.47	.479	1.05	.59	1.96	.375			1.17	.556
	Novelty	1.03	.599	1.66	.436	1.55	.461	0.95	.623			1.46	.482
	Structure	9.76	.008**	1.94	.379	1.13	.569	0.09	.957			2.91	.233
	Text quality	5.41	.067	0.65	.724	1.10	.576	1.10	.576			1.31	.52
	References	0.40	.818	1.47	.479	2.90	.235	2.25	.324			3.14	.208
	Timeliness	0.61	.737	1.11	.573	0.36	.836	3.56	.169			1.62	.446
Abundant -Empty	Content	1.34	.513	0.40	.82	0.19	.908	1.34	.513			1.24	.538
	Novelty	0.01	.996	1.53	.466	1.32	.516	1.01	.602			0.73	.694
	Structure	1.10	.576	0.44	.802	0.55	.761	0.82	.664			4.90	.086
	Text quality	0.52	.772	0.72	.697	0.32	.854	0.51	.774			4.36	.113
	References	0.76	.683	1.14	.565	0.68	.712	0.88	.643			2.13	.344
	Timeliness	1.87	.393	3.80	.149	0.17	.917	0.28	.868			0.78	.678
High-Low	Content	2.22	.329	6.20	.045*	0.76	.683	2.55	.28			2.48	.289
	Novelty	3.19	.203	1.65	.438	0.92	.631	1.87	.393			1.43	.489
	Structure	0.86	.65	0.63	.729	0.97	.615	1.46	.482			6.94	.031*
	Text quality	1.11	.574	0.33	.847	0.68	.712	1.58	.455			3.08	.214
	References	3.16	.206	15.94	< .001***	40.52	< .001***	35.68	< .001***	26.90	< .001***	9.548	.008**
	Timeliness	0.73	.693	4.87	.088	1.50	.472	3.52	.172			2.40	.301
Consructive -Negative	Content	5.26	.072	19.16	< .001***	40.79	< .001***	44.26	< .001***	32.26	< .001***	15.84	< .001***
	Novelty	1.79	.409	19.14	< .001***	30.50	< .001***	34.79	< .001***	22.36	< .001***	7.79	.020*
	Structure	2.97	.226	20.63	< .001***	43.60	< .001***	34.34	< .001***	24.57	< .001***	11.20	.004**
	Text quality	2.77	.25	18.26	< .001***	40.52	< .001***	31.23	< .001***	26.90	< .001***	9.55	.008**
	References	3.16	.206	15.94	< .001***	40.52	< .001***	35.68	< .001***	26.90	< .001***	9.548	.008**
	Timeliness	2.00	.368	19.14	< .001***	34.79	< .001***	25.79	< .001***	21.64	< .001***	10.57	.005**

3.2 RQ2: Are Feedback Patterns Related to Rubric Scores?

For version 1, the statistical relationships between review patterns and category scores were similar for both teacher raters (see Tables 6 and 7). Most scores were correlated with the *Constructive-Negative* dimension on most categories and *High-Low* dimension on references. The main difference was teacher B's content score (M = 1.6, SD = 0.7) that was statistically related only to the *Descriptive-Prescriptive* dimension regarding content ($X^2(2) = 6.305$, $p = .043$, 24/14 *Descriptive/Prescriptive* comments) and structure ($X^2(2) = 9.763$, $p = .008$, 24/11 *Prescriptive*/*Descriptive*). That is, content score was related to student reviewers' more descriptive comments regarding content as well as more prescriptive comments concerning structure. It seems natural that descriptive remarks regarding content-related issues were related to content performance score, but peers' remarks concerning structural aspect were also related to content performance. This could be explained by the relatedness of aspects of content and structure in writing as structural delivery has to be based on content material. Comparatively, it also appears that it may be relatively more difficult for peers to provide prescriptive input (indicate steps for improvement) regarding content but not difficult to give prescriptive insight concerning structural aspect. The correlation results also showed that content score was strongly correlated with structure score. Additionally, B's novelty score was related to *High-Low* on content (M = 1.8, SD = 0.5; $X^2(2) = 6.198$, $p = .045$, 28/10 *Low/High*), indicating that novelty score was also related to student reviewer's more low level (than high level) comments regarding content. As content score and novelty score were among lower scores of all six rubrics for both raters, these two aspects were probably more difficult for novice research students to tackle. Also, novelty indeed concerns issues related to content in that novel perspectives have to be derived from the content synthesized that student writers gathered from the literature they found. Teacher A's timeliness score was related to *Empty-Abundant* on structure (M = 2.6, SD = 0.7, $X^2(3) = 14.893$, $p = .002$, 32/3 *Empty*/*Abundant*), indicating that the timeliness score was also related to mostly uncritical comments regarding structure. As the timeliness score and the structure score were among the higher scores for both raters, performances regarding timeliness and structure were comparatively more satisfactory than the other aspects, and hence student reviewers gave uncritical comments concerning structural issues. Although peer reviewers' remarks regarding timeliness aspect were not statistically related to timeliness score, close inspections revealed that peers also gave mostly uncritical comments concerning timeliness.

Table 8. Interaction between the rubrics and the review pattern dimensions. Contingency table results for rater A, manuscript version 2

		Content		Novelty		Structure		Text quality		References		Timeliness	
		X^2	p	X^2	p	X^2	p	X^2	p	X^2	p	X^2	p
Descriptive	Content			5.54	.136	4.51	.105	2.65	.266	5.17	.075	0.70	.704
-Prescriptive	Novelty							0.92	.632			0.19	.907
	Structure			6.40	.094	0.97	.615	4.72	.094	1.97	.373	0.848	.654
	Text quality			2.21	.530	1.45	.485	2.307	.316	2.33	.312	4.99	.082
	References			5.47	.141	1.40	.496	0.41	.814	1.44	.488	0.37	.833
	Timeliness			0.51	.916	1.02	.6	1.10	.577	0.30	.862	4.63	.099
Abundant	Content			3.90	.273	2.84	.242	3.03	.22	2.33	.312	1.09	.579
-Empty	Novelty							3.06	.216			3.65	.161
	Structure			4.38	.224	2.92	.232	1.82	.402	0.66	.719	7.44	.024*
	Text quality			0.55	.907	1.02	.601	0.98	.613	0.22	.897	1.41	.493
	References			2.31	.511	0.40	.819	2.02	.365	2.12	.346	5.92	.052
	Timeliness			1.87	.601	0.49	.782	1.60	.449	1.87	.393	6.22	.045*
High	Content			2.46	.483	6.59	.037*	4.99	.082	0.77	.68	2.34	.31
-Low	Novelty							4.92	.085			3.53	.172
	Structure			3.51	.320	3.69	.158	2.47	.291	0.68	.711	10.17	.006**
	Text quality			1.19	.757	2.19	.335	0.12	.941	0.86	.651	0.70	.703
	References	13.00	.005**	11.97	.007**	15.94	< .001***	2.39	.303	10.52	.005**	19.806	< .001***
	Timeliness			2.81	.422	3.34	.188	0.70	.704	3.40	.183	4.67	.097
Constructive	Content	14.42	.002**	16.95	< .001***	16.00	< .001***	6.53	.038*	14.90	< .001***	27.84	< .001***
-Negative	Novelty	14.00	.003**	18.00	< .001***	21.71	< .001***	2.64	.267	15.93	< .001***	21.93	< .001***
	Structure	16.54	< .001***	13.34	.004**	21.66	< .001***	2.11	.347	12.06	.002**	22.86	< .001***
	Text quality	13.52	.004**	12.48	.006**	18.26	< .001***	1.03	.597	11.29	.004**	16.71	< .001***
	References	13.00	.005**	11.97	.007**	15.94	< .001***	2.39	.303	10.52	.005**	19.81	< .001***
	Timeliness	12.86	.005**	12.29	.006**	16.36	< .001***	0.50	.779	10.57	.005**	15.07	< .001***

Table 9. Interaction between the rubrics and the review pattern dimensions. Contingency table results for rater B, manuscript version 2

		Content		Novelty		Structure		Text quality		References		Timeliness	
		X^2	p	X^2	p	X^2	p	X^2	p	X^2	p	X^2	p
Descriptive -Prescriptive	Content	6.22	.183	6.06	.048*	1.06	.304	2.79	.594	4.49	.344	4.59	.332
	Novelty			2.54	.281	1.46	.227	2.20	.699	3.66	.455		
	Structure	10.51	.033*	1.96	.375	0.03	.861	10.45	.034*	7.58	.108		
	Text quality			0.73	.696	0.02	.902	2.25	.69	2.76	.599	17.23	.002**
	References			0.13	.936	0.47	.492			4.77	.311		
	Timeliness			3.88	.144	0.19	.662	13.93	.008**	17.95	< .001***		
Abundant -Empty	Content	2.79	.594	1.63	.443	0.73	.392	1.99	.738	2.72	.605	13.20	.010**
	Novelty			0.55	.759	0.22	.64	1.50	.827	1.22	.875		
	Structure	3.10	.541	0.39	.824	1.53	.217	3.74	.443	2.80	.591		
	Text quality			0.22	.897	1.18	.278	1.76	.78	2.96	.565		
	References			0.62	.733	0.11	.739			2.96	.564		
	Timeliness			1.87	.393	0.58	.448	0.90	.925	8.64	.071		
High -Low	Content	4.52	.341	8.13	.017*	0.84	.359	1.71	.788	6.33	.176	10.68	.030*
	Novelty			2.95	.229	1.00	.318	3.55	.47	3.42	.491		
	Structure	2.79	.594	1.67	.433	1.71	.191	6.49	.166	4.19	.381		
	Text quality			0.47	.792	2.53	.112	3.08	.544	6.36	.174		
	References	26.90	< .001***	12.45	.002**	3.90	.048*	28.84	< .001***	17.23	.002**	17.23	.002**
	Timeliness			3.85	.146	0.62	.43	1.73	.785	3.47	.483		
Constructive -Negative	Content	25.16	< .001***	17.42	< .001***	5.16	.023*	42.53	< .001***	13.84	.008**	14.90	.005**
	Novelty	34.50	< .001***	12.07	.002**	5.14	.023*	29.86	< .001***	10.21	.037*	13.79	.008**
	Structure	24.00	< .001***	11.20	.004**	3.46	.063	34.00	< .001***	10.57	.032*	16.86	.002**
	Text quality	35.94	< .001***	11.29	.004**	2.61	.106	26.58	< .001***	15.94	.003**		
	References	26.90	< .001***	12.45	.002**	3.90	.048*	28.84	< .001***	17.23	.002**	17.23	.002**
	Timeliness	27.36	< .001***	10.57	.005**	2.29	.131	22.00	< .001***	11.64	.020*	16.64	.002**

The results of version 2 (see Tables 8 and 9) exhibited more statistically significant relations than version 1 for both teacher raters. The *Empty-Abundant* dimension was significantly related to the timeliness score: for A (M = 2.6, SD = 0.7), on structure ($X^2(2) = 7.438$, p = .024, 32/3 *Empty*/*Abundant*) and timeliness ($X^2(2) = 6.222$, p = .045, 27/1 *Empty*/*Abundant*); for B (M = 3.5, SD = 1.3), on content ($X^2(4) = 13.199$, p = .01, 32/6 *Empty*/*Abundant*). That is, uncritical comments regarding structure and timeliness were correlated with A's timeliness score while uncritical comments concerning content category were correlated with B's timeliness score. Student reviewers' uncritical comments regarding structure category were related to rater A's timeliness score as in version 1; in addition, timeliness score was also directly related to uncritical comments regarding timeliness issue.

More statistically significant relationships were exhibited for rater B's content score (M = 2.5, SD = 1.0) than in B's

version 1 (M =1.6, SD = 0.7). The *Constructive-Negative* dimension on all rubrics (all *Constructive* comments) was all significantly related to the version 2 content score, indicating that all student peers' constructive comments on all writing aspects were reflected in content score. Understandably, content was the main rubric that concerned the ideas presented, and peers tended to be positive rather than negative when giving feedback concerning all aspects. Also, the *Descriptive-Prescriptive* dimension on structure ($X^2(4) = 10.510$, $p = .033$, 24/11 *Prescriptive*/*Descriptive*) and *High-Low* dimension on references ($X^2(4) = 26.903$, $p < .001$, all *Low*) were also significantly related to content score. That is, more prescriptive comments concerning structure and low-level comments regarding references were related to content score. The results showed that student peers tended to give more prescriptive feedback (than descriptive) regarding structural issues, but one would expect peer reviewers to also give prescriptive/descriptive comments regarding content-related issues (as in version 1). Close inspections indicated that student reviewers indeed also provided prescriptive/descriptive remarks concerning content-related aspect, but it was not statistically related. As there existed a strong relationship between content and structure aspects (content score and structure score were correlated) and that structural comments had to be derived from content material displayed, this may explain why student peers' prescriptive/descriptive comments concerning structure were statistically related to content score. The fact that student peers gave low level remarks concerning references issue and also that the references score was the lowest of all rubrics for both raters indicated that the references rubric was challenging both for student writers and student reviewers. The references that a writer finds and employs in the writing would dictate the content that is conveyed; hence, this may explain why peers' low-level comments regarding references were related to content performance.

The exceptions were rater A's text quality score and B's structure score. A's text quality score ($M = 3.2$, $SD = 0.6$) was significantly related to the *Constructive-Negative* dimension on content ($X^2(2) = 6.526$, $p = .038$, all *Constructive*). In other words, text quality score was closely related to student peers' positive comments concerning content category. Since text quality concerns expression clarity and consistency that convey content (content score and text quality score were also found to be moderately related) and that structural delivery is based on content as input, the result was not unexpected. B's structure score ($M = 2.7$, $SD = 0.5$) was closely related to *Constructive-Negative* dimensions on content ($X^2(1) = 5.158$, $p = .023$, all *Constructive*), novelty ($X^2(1) = 5.143$, $p = .023$, all *Constructive*), and references ($X^2(1) = 3.903$, $p = .048$, all *Constructive*). Similarly, as in version 1, the *High-Low* dimension on references was also associated with structure score ($X^2(1) = 3.903$, $p = .048$, all *Low*). That is, structure score was closely related with student peers' *Constructive* comments regarding content, novelty, and references categories as well as *Low* level comments concerning references category. As references category was the common denominator of the two feedback dimensions (constructive comment and low-level comment) that were closely related with structure performance, it may be of interest to know what role the references rubric has to play in this regard. One interpretation could be that references that authors gathered may affect content orientation, which in turn may influence novelty perspectives, and hence also structural delivery.

Rater B had more performance scores associated with the *Descriptive-Prescriptive* dimension than rater A. For assessor B, *Descriptive-Prescriptive* dimension was significantly related to four rubric scores: (a) the content score as mentioned above; (b) the novelty score (on content: $M = 2.7$, $SD = 0.6$, $X^2(2) = 6.063$, $p = .048$, 24/14 *Descriptive*/*Prescriptive*); (c) the text quality score ($M = 2.9$, $SD = 0.8$, on structure: $X^2(4) = 10.446$, $p = .034$; 24/11 *Prescriptive*/*Descriptive*; on timeliness: $X^2(4) = 13.928$, $p = .008$, 26/2 *Descriptive*/*Prescriptive*); and (d) references score (on timeliness: $M = 2.4$, $SD = 1.2$, $X^2(4) = 17.949$, $p = .001$, 26/2 *Descriptive/Prescriptive*). In other words, the content score was related to more *Prescriptive* comments regarding structure, indicating that student peers' providing direction for improvement concerning structural aspect reflected content performance. Close inspection showed that content score was among lower scores of all rubrics, despite having greatly improved from version 1 to version 2. Probably providing structural feedback for student reviewers was comparatively more readily feasible than remarking on content-related issues which may involve certain subject-matter knowledge that not all students had or felt confident enough to share with the writers. Also, structural comments had to be based on content material, and hence this result was interpretable. The novelty score was related to more *Descriptive* comments on content, indicating that novelty performance could be detected from peers' comments describing the state concerning content of the manuscript. Novelty concerns new perspectives derived from the literature that a student has gathered, and thus arguably this may be one of the challenging aspects for both student writers and student reviewers. This may explain why novelty performance was statistically related to student reviewers' comments depicting the current state of content. Text quality score was related with more *Prescriptive* comments on structure and *Descriptive* comments on timeliness, indicating that student reviewers' comments providing improvement steps regarding structural aspect and describing the state concerning timeliness issue were related to text quality performance. As text quality score was strongly correlated with

structural score, and also it may be relatively straightforward to provide suggestions for structural improvement, this result was perhaps not surprising. As both text quality and timeliness were among higher scores of all rubrics, the performance of the two rubrics were comparatively more satisfactory than other rubrics; hence, describing the state of timeliness may be quite straightforward for student reviewers. Also, text quality directly affects readability that may impact a reader's impression concerning topic adequacy (timeliness); this may account for why comments describing timeliness were related to text quality score. References score was related to more *Descriptive* comments on timeliness, indicating that peers described the state of timeliness issue but not pinpointing revision suggestions.

3.3 RQ3: Are students' final grades related to review patterns and review quantity?

No correlation was observed between students' collective final grades and the review dimensions. Also, no significant correlations were found between final grades and the amount of feedback received ($r(21) = .039$, $p = .867$). Hence, there was no relationship between final grades and feedback patterns, nor was there between final grades and review quantity.

3.4 RQ4: Are reviewers' grades related to the review patterns they provide?

The dimension of *High* level comments was the only one to significantly correlate with reviewers' grades ($r(21) = .207$, $p = .04$). No significant relationships were found between reviewers' own grades and the reviewing patterns.

4. Discussion

4.1 RQ1: What Review Patterns Do Writing Categories Exhibit?

There was a distinct trend across the three student reviewers for each manuscript in that the reviewers seem to have had consistent emphasis. The findings suggest that descriptive, superficial or uncritical, low level, and constructive comments were possibly the easiest (or safest) way to provide comments to the peers. To provide more prescriptive insight, student reviewers may have required more time and effort to read, understand, verify, and deliver in a second language (English). Given that time was limited, and each reviewer (who was to give feedback) was also an author (who was to revise accordingly and also respond to feedback), such a pattern may result from students' choosing to simplify the reviewing task. Also observed was that when a reviewer gave both prescriptive and high-level comments, he/she also tended to provide more concrete/critical remarks, suggesting that certain reviewers may have been more mature in terms of content knowledge and text quality (clarity and consistency). The least common pattern was prescriptive, empty, low level, and negative.

For the second round feedback, the reviewers responded to more categories than the first round review. There may be several reasons. The reviewers may have by then have become more accustomed to providing feedback as required though there also were two in-class practice sessions. This may also be interpreted as that the authors have added more material after the first round of review based on peer comments or self-assessment or both (as the authors also were to provide a response memo for the first feedback, and hence self-evaluations were done), and consequently the reviewers have more material to comment upon. Also, the content category appears to have been easier to comment on compared to the other categories, while novelty was the hardest to comment on. Overall, structure, text quality, references, and timeliness were provided with more comments in the second review. The peers' comments on the first version appeared to be more prescriptive in terms of giving advice, while comments on the second version tended to be more descriptive in forms of stating the status. It may be that the first version is often less developed, and thus peer reviewers would tend to prescribe solutions or suggestions compared to the second version in which revision is already done after the first feedback and thus the reviewers simply describe what the author has achieved. Peer comments on version 1 also contain more high-level comments than on version 2, suggesting that when writing was not yet matured, the reviewers tended to give high level remarks (such as on subject knowledge and structure). Reviewers' comments on version 1 also contained more abundant input than for version 2, suggesting that student writers were able to make use of reviewers' input after receiving feedback and thus likely triggering fewer comments in version 2. All the six writing categories exhibited a similar trend in that the reviewers' comments tended to be more descriptive, more low level, superficial, and constructive in version 2.

Most comments were of type constructive, while few were of type negative, suggesting that student reviewers were aware of how their feedback may be perceived and thus were attentive to be constructive as instructed. However, overall empty/superficial or uncritical comments occured more frequently than abundant comments. Several explanations may be possible. One likelihood may be that some student reviewers would in their tendency be minimalist for egoistic reasons since the reviewers' feedback did not count towards the final grade. Another reason may be that giving abundant feedback based on the six criteria to consider the six writing aspects requires time and devotion to properly read and understand each manuscript of a considerable length. A perhaps less likely explanation

may be from the perspectives of students' individual culture where potential criticism should be avoided. In this line of thinking, it is better not to give any comments that may be interpreted in different ways. If one is forced to give comments, then it is safer to have uncritical or superficial comments than to confront writers with real thoughts.

Further, student assessors employed comparatively more low-level comments than high-level comments, which was also observed in previous studies (Cho & MacArthur, 2010; Stevenson, Schoonen, & De Glopper, 2006). Possible explanations have been associated with the students' insufficient skills of target language writing (English as L2) and lacking task schemata knowledge involving accessing prior information learned (Stevenson et al., 2006), and unaware of writing problems as well as lacking knowledge of genre-specific features (Alamargot & Chanquoy, 2001). In the present study, being unaware of genre-specific features such as literature reviews writing was perhaps unlikely since this genre feature is specifically instructed in class as well as students' own reading outside class. The researcher would argue that factors associated with subject knowledge, metacognitive knowledge and experiences, learner efficacy, and learner motivation may be of greater impact than of English proficiency issues, given that these student reviewers are master-level research learners who have had multiple experiences of reading and writing in prior contexts. It is also likely that new research students may have found it weary to provide content-related remarks as at the master-level they were aware that knowledge deliverance was subject to scrutiny, while it was less threatening to offer advice concerning structural issues that are based on content material delivered.

Content and timeliness contained a mixture of low-level and high-level comments. As observed, high-level comments were often accompanied by low-level comments but not the other way around. Low-level comments also often were descriptive (describing what the writing had achieved) and praising that the authors' topics were good or reasonable, particularly for timeliness. The findings suggest that reviewers were able to provide both low-level and high-level remarks for these two categories. Novelty was given more high-level remarks. This is understandable as novelty refers to a high-level abstract concept that the authors synthesised based on the literature obtained. The structure category contained both prescriptive and descriptive comments in both versions, but only in version 1 did it contain high-level comments, indicating a less mature structure in an earlier version would often trigger remarks concerning globally-related issues (content material division and section signposting). When text was mature as in version 2, prescriptive advice regarding structure was also likely as reviewers could see potentially better organization that writers may not see immediately despite mostly being low-level comments (hence prescriptive but low-level). Structure (material organization), text quality (clarity and consistency), and references (sources obtained) appeared to be given more low-level comments. This could be interpreted as that when considering these three categories, most student reviewers tended to provide their feedback in such a way that satisfied the immediate goal of local improvement at the expense of higher-order concerns manifested in the manuscripts that they review, despite that both levels of feedback could be applied and supplied. Further research into reviewer tendency and strategy in tasks involving multiple categories is needed to shed light in this regard. Also, structure and text quality seem closely linked, i.e., if one answers structure, one also tends to answer text quality. The analyses revealed that structure performance and text quality performance are indeed statistically closely related.

The answer to RQ1 is that the six writing categories indeed exhibit distinct feedback patterns. The student assessors were able to provide feedback based on the six writing criteria, designed to guide writers compose their literature reviews writing as well as assisting peer assessors in providing written feedback to their peer writers.

4.2 RQ2: Are Feedback Patterns Related to Rubric Scores?

The constructive vs. negative and high vs. low level review pattern dimensions correlated most strongly with rubric scores for both versions. All peers' constructive vs. negative comments on writing rubrics (content, novelty, structure, text quality, references, and timeliness) were significantly correlated with most rubric scores when rated by two raters, with a few exceptions. This finding suggests that student reviewers provided constructive (instead of negative) comments on all separate categories in their peers' literature survey writing as well as low/high level remarks on references that contributed to their category performance, assuming all student writers addressed all comments raised by their peer reviewers.

In addition, the dimensions of empty/abundant and high/low-level regarding structure category correlated with timeliness score in version 1 rated by rater A, showing that uncritical and low-level remarks regarding structure aspect also reflect timeliness performance. Similarly, the high/low-level dimension on structure category correlated with version 1 timeliness performance when scored by rater B. Version 1 timeliness performance and structure performance were among the higher scores of all categories, indicating that the two categories were more satisfactory compared to the other aspects in version 1. One explanation may be that the two aspects were comparatively easier to achieve for student writers and also easier for student reviewers to provide feedback in terms of depth dimension and

level dimension (although mostly superficial and most low level than high. This may also suggest that superficial/uncritical comments may suggest at least satisfactory performance, if not top performance, whereas more likely high/low level remarks may apply to performance of all kinds since all manuscripts may benefit from both high or level feedback.

More statistically significant relationships between peer feedback patterns and teacher category scores were found for version 2 than version 1. The high/low-level dimension regarding content category correlated with the structure score by rater A (highest of all rubrics by A) and also with the novelty score (third rank of six rubrics by B) and timeliness score by rater B (highest of all rubrics by B), with mostly low-level remarks, suggesting more low-level comments regarding content category were linked with scores of structure, novelty, and timeliness. Put differently, student reviewers' feedback concerning content category, even if being mostly low-level, may have reflected or contributed to writing performance of structure, novelty, and timeliness. This is explainable as content material revealed topic adequacy (timeliness) and original perspectives (novelty) derived from the literature gathered and it affected overall material organization (structure). The empty/abundant dimension regarding timeliness category and structure category (by rater A) as well as content category (by rater B) were associated with timeliness score, suggesting that peer reviewers' uncritical/superficial comments regarding timeliness category, structure category, and content category reflected the version 2 timeliness score. Version 2 timeliness performance and structure performance were also among the higher scores by both raters while content performance was ranked as mid-to-lower scores (by A as mid and B as fifth of six rubrics). It is hard to explain why superficial/uncritical comments were related to the lower content score, if one would accept that uncritical comments may reflect at least satisfactory performance but not reflect lower then middle performance. One possible explanation may be that content domain with specific subject knowledge made it harder to offer in-depth comments for student reviewers; this was also displayed from the mid-to-lower scores in student authors' content performance. Still, the trend was that higher scores were associated with more abundant comments (less superficial comments) than lower scores, but comparatively much fewer abundant remarks were given. In other words, more abundant feedback may benefit writers more than superficial comments do as reflected in the category scores.

The descriptive/prescriptive dimension concerning structure category correlated with the content score and text quality score and also the descriptive/prescriptive dimension regarding timeliness category correlates with text quality score and references score when rated by rater B. The student reviewers provided twice as many prescriptive as descriptive remarks onthe structural aspect, offering more advice or suggestions than simply describing what the authors have done; this comment dimension was linked to writers' content performance and text quality performance that were statistically strongly correlated. The statistical data also revealed that structure performance was also strongly correlated with text quality performance. It is thus explainable that student reviewers' descriptive/prescriptive dimension regarding the structural aspect may have been linked to student authors' content score and text quality score. Similarly, the student reviewers offered descriptive/prescriptive feedback dimension regarding the timeliness category but provided more confirming remarks than give advice or suggestions, indicating that the timeliness issue was already satisfactory, which was confirmed as its top score (of all rubrics). The fact that the student reviewers' feedback concerning timeliness corresponds to student authors' text quality performance and references performance reveals that reviewers' describing topic adequacy (timeliness) being good is related to authors' text clarity and consistency (text quality), suggesting that readability was linked to overall impression of topic relevance and that topic relevance as well as recency (timeliness) was also linked to references that authors gathered.

The findings are consistent with those of previous studies in that student reviewers were able to provide useful feedback for their peers' writing (Snowball & Mostert, 2013; Graham & Perin, 2007). The findings also suggest that most category scores may be predicted by referring to student reviewers' constructive vs. negative comments on all writing categories and high vs. low level comments on references category. Particularly, being able to produce quality references was one important component for writing successful literature surveys because quality references affect content construction, original perspectives (novelty), structure that was based on content conveyed, and topic appropriateness (timeliness).

The answer to RQ2 is that certain review patterns were related to certain writing category scores; in particular, constructive comments on nearly all categories and more low-level comments on references category were related to most category performances. The findings may be interpreted as that student reviewers' certain written feedback patterns reflected student writers' category performances even when rated/scored by teacher raters based on the same six categories, indicating that student reviewers' feedback assists their peers' writing in addition to student writers' possible self-evaluation and self-revisions over the writing process as well as the peer feedback process.

4.3 RQ3: Are Students' Final Grades Related to Review Patterns and Review Quantity?

No correlations were found between final grades and review patterns received; also, no correlations were detected between the student writers' final grades and the amount of feedback received. The findings agree with the previous studies in that peer feedback quality was not related to writing performance but that all writers benefit from peer feedback concerning the different writing aspects of essay text (Huisman et al., 2017). In the present study all student authors improved from version 1 to version 2 in all separate category scores, having received two opportunities of peer written feedback on their literature reviews writing genre. However, the authors' final grade was not related to peers' review comments. The present study also agrees with the finding that participating in peer feedback did not significantly define performance mark (Snowball & Mostert, 2013). The present findings also appear to be different from the previous studies (Cho & Schunn, 2007; Karegianes, Pascarella, & Pflaum, 1980) where the greater number of peer feedback induces the more improvement. It may be that the present study reflects that a collective grade is harder to associate with criteria assessment based on six writing components (hence six category scores). It is possible that one or more categories were weighed more importantly when determining a final grade. The same phenomenon has also been reported in previous studies where a global measure of performance could have cluttered the effect of individual categories (Kim, 2005; Sluijsmans et al., 2002; Prins, Sluijsmans, & Kirschner, 2006). Further studies are needed to ascertain what determines a final grade and how to derive it based on the six writing components. The fact that the final grade and the amount of feedback received were not correlated suggests that feedback quantity was unlikely to be the decisive factor that contributed to a single performance grade. Other factors could also affect writing performance, such as degree of authors' own effort and writing proficiency, learner efficacy, learner attitudes, authors' own subject-matter knowledge, reviewers' disciplinary knowledge, and resources including library subscription. It is also worth noting that review accuracy was less important than expected, which is in agreement with the previous finding (Gielen, Peeters, et al., 2010). Thus, the answer to RQ3 is that student manuscripts' final performance mark was not related to review patterns received, nor was it related to the quantity of reviews received.

4.4 RQ4: Are Reviewers' Grades Related to the Review Patterns They Provide?

Only high-level (H) review dimension correlated with reviewers' own performance grade, suggesting that a student reviewer's ability demonstrated in his or her manuscript writing was not related to the feedback he or she provided for peer students, except the type of high-level comment. These student reviewers often provided high-level comments along with low-level comments, hence also providing more abundant than superficial comments. The findings thus suggest that the ability to provide high-level comment is one predictor of student assessors' literature reviews writing performance. Similar observations were also confirmed in that high-achieving peer assessors provide more content-related feedback although individual ability did not define feedback quality, and that feedback quality was not related to writing performance (Huisman et al., 2017). The answer to RQ4 is that only high-level feedback type was related to the performance grades of the reviewers.

5. Limitations and Future Work

The sample for this study was small (n = 21). Future studies should include larger sample sizes to enhance the generalizability and validity. The study solely explored student reviewers' written feedback type (what they provide as feedback was coded as comment/dimension type) based on six writing categories/assessment criteria designed for literature reviews genre with reference to writing performance. The student reviewers were also writers who submitted their manuscripts to three student assessors. The assumptions were that all authors addressed all comments raised by all reviewers. This, however, is a practical but simplistic view as authors improve their writing constantly and not all elements may be measured or quantified. A different design would be needed to specifically measure what authors have improved based on the comments received and what have improved but not based on the comments received as well as what have not improved based on the comments received. At a more advanced level as a literature reviews writing for new master students, aspects such as novelty, timeliness, and references would be vital in terms of producing newness of knowledge. A useful approach would probably be asking student reviewers to keep a journal log of what are of challenging areas upon which to provide feedback and by requesting student authors to keep track of all their thoughts during revision stage and all individual learning stages that different authors may employ. Further methodological design based on such cross examinations may be a feasible direction. Future work could also further investigate author responses to reviewers' remarks to scrutinise other factors contributing to performance improvement.

6. Conclusion

This study explored first-year master student assessors' feedback pattern characteristics in assessing literature

reviews genre based on six assessment criteria on the same six writing categories. The findings indicated statistically significant relationships between specific peer feedback patterns and student authors' categorized scores in the six writing aspects in both revisions completed during two peer review processes. The constructive/negative dimension correlated with several of the separate category performance along with the high/low-level dimension, particularly with regards to the references category. The most common peer feedback patterns were descriptive (describing the state but offering no advice), superficial/uncritical, low-level, and constructive while the least common was prescriptive (prescribing directions for revision), superficial/uncritical, low-level, and negative. No correlation was found between beginner-level research students' final collective grade, review characteristics, and the quantity of feedback received. Overall, the findings indicate that peer feedback quality and quantity did not significantly determine the literature reviews writing performance as collective mark; however, individual writing components of literature reviews genre showed improvement between revision drafts over the writing process and over the peer feedback process.

References

Alamargot, D., & Chanquoy, L. (2001). *Through the models of writing*. Dordrecht, the Netherlands: Kluwer. https://doi.org/10.1007/978-94-010-0804-4

American Psychological Association. (2009). *The Publication Manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Anderson, T., Alexander, I., & Saunders, G. (2020). An examination of education-based dissertation macrostructures. *Journal of English for Academic Purposes, 45*, 100845. https://doi.org/10.1016/j.jeap.2020.100845

Basturkmen, H. (2009). Commenting on results in published research articles and masters dissertations in Language Teaching. *Journal of English for Academic Purposes, 8*(4), 241-251. https://doi.org/10.1016/j.jeap.2009.07.001

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn* (Vol. 11). Washington, DC: National academy press.

Bunton, D. (2005). The structure of PhD conclusion chapters. *Journal of English for Academic Purposes, 4*(3), 207-224. https://doi.org/10.1016/j.jeap.2005.03.004

Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction, 20*(4), 328-338. https://doi.org/10.1016/j.learninstruc.2009.08.006

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education, 48*(3), 409-426. https://doi.org/10.1016/j.compedu.2005.02.00

DiPardo, A., & Freedman, S. W. (1988). Peer response groups in the writing classroom: Theoretic foundations and new directions. *Review of Educational Research, 58*(2), 119-149. https://doi.org/10.2307/1170332

El-Dakhs, D. A. S. (2018). Why are abstracts in PhD theses and research articles different? A genre-specific perspective. *Journal of English for Academic Purposes, 36*, 48-60. https://doi.org/10.1016/j.jeap.2018.09.005

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist, 34*(10), 906. https://doi.org/10.1037/0003-066X.34.10.906

Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, Diagnosis, and the Strategies of Revision +. Composition. *College Composition and Communication, 37*(1), 16-55. https://doi.org/10.2307/357381

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20*(4), 304-315. https://doi.org/10.1016/j.learninstruc.2009.08.007

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*, 445-476. https://doi.org/10.1037/0022-0663.99.3.445

Hanrahan, S. J., & Isaacs, G. (2001). Assessing self-and peer-assessment: The students' views. *Higher Education Research & Development, 20*(1), 53-70. https://doi.org/10.1080/07294360123776

Hopkins, A., & Dudley-Evans, T. (1988). A genre-based investigation of the discussion sections in articles and dissertations. *English for Specific Purposes, 7*(2), 113-121. https://doi.org/10.1016/0889-4906(88)90029-4

Huisman, B., Saab, N., van den Broek, P., & van Driel, J. (2019). The impact of formative peer feedback on higher education students' academic writing: a Meta-Analysis. *Assessment & Evaluation in Higher Education, 44*(6), 863-880. https://doi.org/10.1080/02602938.2018.1545896

Huisman, B., Saab, N., van Driel, J., & van den Broek, P. (2017). Peer feedback on college students' writing: exploring the relation between students' ability match, feedback quality and essay performance. *Higher Education Research & Development, 36*(7), 1433-1447. https://doi.org/10.1080/07294360.2017.1325854

Karegianes, M. L., Pascarella, E. T., & Pflaum, S. W. (1980). The effects of peer editing on the writing proficiency of low-achieving ten-grade students. *Journal of Educational Research, 73*, 203-207. https://doi.org/10.1080/00220671.1980.10885236

Kawase, T. (2015). Metadiscourse in the introductions of PhD theses and research articles. *Journal of English for Academic Purposes, 20*, 114-124. https://doi.org/10.1016/j.jeap.2015.08.006

Khedri, M., Heng, C. S., & Ebrahimi, S. F. (2013). An exploration of interactive metadiscourse markers in academic research article abstracts in two disciplines. *Discourse Studies, 15*(3), 319-331. https://doi.org/10.1177/1461445613480588

Kim, M. (2005). *The effects of the assessor and assessee's roles on preservice teachers' metacognitive awareness, performance, and attitude in a technology-related design task*. Unpublished doctoral dissertation, Florida State University, Tallahassee, USA.

McConlogue, T. (2015). Making judgements: investigating the process of composing and receiving peer feedback. *Studies in Higher education, 40*(9), 1495-1506. https://doi.org/10.1080/03075079.2013.868878

Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: how students respond to peers' texts of varying quality. *Instructional Science, 43*(5), 591-614. https://doi.org/10.1007/s11251-015-9353-x

Phillipson, R. (2012). Linguistic imperialism. *The Encyclopaedia of Applied Linguistics*, 1-7. https://doi.org/10.1002/9781405198431.wbeal0718.pub2

Prins, F., Sluijsmans, D., & Kirschner, P. A. (2006). Feedback for general practitioners in training: quality, styles, and preferences. *Advances in Health Sciences Education, 11*, 289-303. https://doi.org/10.1007/s10459-005-3250-z

Rijlaarsdam, G., & Couzijn, M. (2000). What do writers learn from peer comments on argumentative texts? In A. Camps, & M. Milian (Eds.), *Metalinguistic activity in learning to write*. In Rijlaarsdam, G., & Esperet, E. (Eds.), *Studies in Writing, Vol. 6* (pp. 167-202). Amsterdam: Amsterdam University Press.

Sluijsmans, D. M. A., Brand-Gruwel, S., & Van Merrienboer, J. J. G. (2002). Peer assessment training in teacher education: effects on performance and perceptions. *Assessment & Evaluation in Higher Education, 27*, 443-454. https://doi.org/10.1080/0260293022000009311

Snowball, J. D., & Mostert, M. (2013). Dancing with the devil: Formative peer assessment and academic performance. *Higher Education Research & Development, 32*(4), 646-659. https://doi.org/10.1080/07294360.2012.705262

Stanley, J. (1992). Coaching student writers to be effective peer evaluators. *Journal of Second Language Writing, 1*, 217-233. https://doi.org/10.1016/1060-3743(92)90004-9

Stellmack, M. A., Keenan, N. K., Sandidge, R. R., Sippl, A. L., & Konheim-Kalkstein, Y. L. (2012). Review, revise, and resubmit: The effects of self-critique, peer review, and instructor feedback on student writing. *Teaching of Psychology, 39*(4), 235-244. https://doi.org/10.1177/0098628312456589

Stevenson, M., Schoonen, R., & De Glopper, K. (2006). Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing, 15*, 201-233. https://doi.org/10.1016/j.jslw.2006.06.002

Tang, G., & Thitecott, J. (1999). Peer response in ESL writing. *TESL Canada Journal*, *16*(2), 20-38. https://doi.org/10.18806/tesl.v16i2.716

Tankó, G. (2017). Literary research article abstracts: An analysis of rhetorical moves and their linguistic realizations. *Journal of English for Academic Purposes, 27*, 42-55. https://doi.org/10.1016/j.jeap.2017.04.003

Van Steendam, E., Rijlaarsdam, G.C.W., Sercu, L., & Van den Bergh, H.H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction, 20*(4), 316-327. https://doi.org/10.1016/j.learninstruc.2009.08.009

Van Zundert, M., Sluijsmans, D., & Van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*(4), 270-279.

https://doi.org/10.1016/j.learninstruc.2009.08.004

Vickerman, P. (2009). Student perspectives on formative peer assessment: an attempt to deepen learning? *Assessment & Evaluation in Higher Education, 34*(2), 221-230. https://doi.org/10.1080/02602930801955986

Vygotsky, L. S. (1986). *Thought and language* (Revised edition). Cambridge, MA: Massachusetts Institute of Technology.

Xie, S. (2020). Multidimensional analysis of Master thesis abstracts: a diachronic perspective. *Scientometrics, 123*, 861-881. https://doi.org/10.1007/s11192-020-03408-6