

**LITERATURE REVIEW WRITING AS A LEARNER-CENTRIC PROCESS
BY BEGINNER-LEVEL RESEARCH STUDENTS: HOW EFFECTIVE ARE
RUBRICS FOR ASSESSING STUDENTS' WRITING?**

Evelyn Eika

*Faculty of Technology, Art and Design, Oslo Metropolitan University, P.O. Box 4 St.
Olavs plass, 0130 Oslo, Norway
Email: Evelyn.Eika@oslomet.no*

ABSTRACT: *Developing academic writing skills is an essential part of most advanced-level degrees. Writing can help a student identify structure within a complex domain, organise thoughts, and communicate ideas. This study explores literature review writing by beginner-level research students as their beginning steps to approach research writing. A controlled study involving a class of master students writing literature surveys was designed. Six rubrics (content, structure, text quality, novelty, timeliness, and references) were proposed to help guide students' writing process accompanied by peer feedback sessions and to assist assessors' assessing the writing. The results revealed that the use of the six rubrics was beneficial for students' writing process and valid for assessors' assessment. Peer feedback showed a positive impact on writing performance. The references rubric was observed as the most effective predictor of grade that may serve as a motivator to engage students in revising for improvement.*

KEYWORDS: Academic English writing, Higher education, Research students, Literature reviews as text genre, Rubrics, Peer feedback

INTRODUCTION

Literature Reviews as Text Type

English today is regarded as a world language as it is widely used in international contexts for various communicative purposes (Bhatia, 2019). As the global spread of English use increases, publishing in English has become a trend for educators and scholars to disseminate their work in the global arena. Among scholarly works based on English for specific purposes or English for academic purposes, a vast amount of research has focused on published research articles and students' degree dissertations. Most of these works dealt with the individual sections of research articles or dissertations in various disciplines, including abstracts (e.g., Xie, 2020; Tanko, 2017; El-Dakhs, 2018; Khedri et al., 2013; Hartley, 2003), introductions (e.g., Kawase, 2018), results (Basturkmen, 2009), discussions (Hopkins & Dudley-Evans, 1988), and conclusions (Bunton, 2005). One recent research has also focused on macrostructures of doctoral dissertations (Anderson et al., 2020). There appear to be few studies on literature reviews as text type or genre. According to APA (The American Psychological Association, 2019, 2009), a literature review is an article type which

serves as a tutorial for a chosen topic area by defining the problem, describing recent relevant research, and suggesting future research steps:

Literature reviews, including research syntheses and meta-analyses, are critical evaluations of material that has already been published.[...] By organizing, integrating, and evaluating previously published material, authors of literature reviews consider the progress of research toward clarifying a problem. In a sense, literature reviews are tutorials, in that authors define and clarify the problem, summarize previous investigations to inform the reader of the state of research, identify relations, contradictions, gaps, and inconsistencies in the literature, and suggest the next step or steps in solving the problem. (2009, p. 10)

Concerning the structure, the APA states that “the components of literature reviews can be arranged in various ways (e.g., by grouping research based on similarity in the concepts or theories of interest, methodological similarities among the studies reviewed, or the historical development of the field)” (2009, p. 10). Undoubtedly all research work contains its literature review element, most likely at the beginning sections. Most research articles and dissertations consist of sections of introduction, method, results, and discussion (IMRD), along with abstracts and conclusions. The present study focuses on literature reviews as a text genre that differs from a research article genre reporting one’s own study.

Genre-Based Approach

Genre-based approach has been used to analyse text genres in terms of language features, rhetorical devices, and structure (van Enk & Power, 2017; Swales, 1990). Genre is considered as a goal-oriented social process (Martin, 2009), as social practice employing language (Hyland, 2002), with an emphasis regarding context versus text and focusing on communicative values (Bhatia, 2019). As indicated (Hyland, 2005), text ought to be viewed holistically, not as separate ideational (propositional), interpersonal, and textual functions, since writers create content to engage readers as a process. The present study explores literature reviews as a text genre and as a writing process by engaging student writers through stages of writing tasks.

Globalisation and Students’ Writing

Possibly due to the trend of increased globalisation, the number of students enrolled in higher education has greatly increased in countries such as Canada (Anderson, Alexander, & Saunders, 2020) and Norway (Nygård, 2020). Within a period of 17 years (2000-2016), Canadian university students increased 55 percent and that of doctoral students increased nearly 100 percent. In Norway, despite a much smaller population, the number of graduations increased about 38 percent within the same period and that of doctoral students increased about 49 percent. Considering the steady increase of bachelor and research students (including master and PhD level) and the increased amount of dissertations that would be produced, it is important to investigate the students’ writing of literature reviews genre as it is the starting point of all research work. Literature review writing can help students understand the existing research

discourse and get inspiration for their own research designs that are realistic and feasible, novel, and meaningful.

Peer Feedback in Writing

Peer assistance in writing has been found to be mostly constructive. For example, learners from fourth-graders to high school students assisting each other in one or more aspects of writing increased their writing quality, more than students working individually (Graham & Perin, 2007). Providing feedback to peer writers also helped students become more aware of audiences (Mendonca & Johnson, 1994). Further, comments provided by single peers were found to be as effective as by teachers (Cho & Schunn, 2007; Gielen et al., 2010). Employing more peer reviewers was even more effective than teacher comments, contributing to improved writing performance (Cho & Schunn, 2007; Karegianes et al., 1980). Furthermore, writers who solely gave feedback made more gains in writing ability compared to those who solely received feedback (Lundstrom & Baker, 2009). Peer feedback was also reported to be beneficial to students in higher education, and that the students could perceive positive gains from peer assessment (Hanrahan & Isaacs, 2001). However, in one study no performance improvement was found in students' revisions, and it was reasoned to be students' inability to internalise peers' comments (Kim, 2005). Concerning the relationship between learner ability and writing performance, Huisman and colleagues (2017) found that students' individual ability was not directly related to essay writing performance; writers of different ability-levels benefited from peer feedback to a similar extent in different writing aspects. Contrastively, Snowball and Mostert (2013) found that participating in peer feedback sessions did not significantly impact the final essay grade but subject-specific ability and English proficiency did.

Writing Assessment Criteria

To help assess students' writing, suitable assessment criteria would be required. Educational Testing Service derived a five-factor rubric (Diederich, French, & Carlton, 1961): ideas (relevance, clarity, quantity, development, and persuasiveness); form (organisation and analysis); flavour (style, interest, and sincerity); mechanics (specific errors in punctuation and grammar); and wording (choice and arrangement of words). This approach has insufficiencies in terms of validity and context since teachers' rhetoric values are often involved while students' values may be left out (Broad, 2003; Turley & Gallagher, 2008).

More recent approaches to writing assessment have turned towards discipline-specific evaluations. For instance, a six-category rubric was developed to assess social work assignments of an undergraduate class (Hooper & Butler, 2008): APA format; punctuation, grammar, spelling, and editing; following provided outline; including relevant quotes from the text; demonstrating insight into personal biases; and identifying application to the field of social work. A four-point holistic rubric was designed to help teachers evaluate secondary school students' literary writing focusing on content as opposed to grammar (Aguirre-Muñoz & Boscardin, 2008): focusing on important character features, supporting their assertions, presenting a coherent

response, and containing mechanical errors. Another study assessing social work students' writing abilities at graduate level employed a six-scale rubric (Alter & Adkins, 2006): diction, sentences, paragraphs, essay, mechanics, and usage. Within an essay, four further categories were created: clear thinking, logical structure, coherence forming logical links, and details as evidence.

The Present Study

Very few studies have investigated writing assessment of literature reviews as text genre by novice research students. This study thus aims to fill this knowledge gap in writing research via a writing process combined with peer feedback. The literature review genre introduces the specific problem under investigation, describes relevant studies and research strategies in logical continuity between works, and develops the problem with breadth and clarity for a professional audience (APA, 2019, 2009; Webster & Watson, 2002). It is stressed that relevant scholarship should contain summaries of the most recent and directly related studies as well as recognising the primacy of others' work. Hence, in addition to employing traditional assessment rubrics such as content, structure, and language (text quality), writing a literature review article requires elements that express newness. The present study employs three additional criteria to help address this issue, i.e., novelty (of perspectives), timeliness (of topic), and references (of recency). Further, it would also be beneficial for both students and educators if any of these criteria would help predict performance level from the perspectives of providing early remedial support for students' learning outcome. In order to gain insight into novice research students' writing of literature reviews as a text genre and as a process, the following research questions were raised:

RQ1: To what degree do the six assessment criteria (content, structure, text quality, novelty, timeliness, and references) agree with the characteristics of the actual literature review writing?

RQ2: To what degree does the writing process involving peer feedback help improve students' writing?

RQ3: Do any of the rubrics predict performance grade?

METHOD

Review Criteria

Rubrics of six writing categories were created to help student writers and peer reviewers focus on specific aspects of the writing relevant to the literature survey genre: content, degree of novelty or originality, structure, text quality, references, and timeliness. Content concerns the clarity of objectives, basis of theory, conclusions drawn, and understanding of the subject. Degree of novelty or originality deals with presence of new ideas and original perspectives on existing literature. Structure concerns organisation of the material, choice of section headings, general layout, and use of figures/diagrams. Quality of text focuses on clarity of expression, consistency, and readability. Timeliness of topic concerns whether the student is addressing a topic that

is relevant today and others are also working on currently. References are a vital part of scientific research, and hence it concerns whether referred previous work is of sufficient quality, quantity, and up-to-date. The student reviewers are also encouraged to do a literature search to corroborate the reference list for each writing they are assigned.

Participants

Twenty-one ($n = 21$) first-year master students studying a computer-related degree participated in the study. Nineteen students were male and two students were female. Their age ranged from 23 to 40. The students used English as a second language and were of varied linguistic backgrounds.

Procedure

This study implemented writing as a process where activities, stages, sociocultural interactions, instructions, and self-reflections were arranged to assist student writers (cf. Pritchard & Honeycutt, 2006).

Authors' writing tasks: The students were asked to submit the first draft of a literature survey of their chosen topic related to information technology of no more than 3,000 words excluding references. Altogether three drafts were required with two peer feedback sessions.

Review task one: The student peer reviewers were then randomly assigned as a form of writing training (cf. Lupton, 2008) to engage students as evaluators in comparing and assessing their peers' work, which in turn would strengthen their self-assessment and improve their own work. Each reviewer had to review three writings. The criteria for reviewing were provided based on the categories described above. To help peer students execute the review task, a review form containing the six writing categories was constructed as a facilitating instrument. Peer reviewers were instructed to be constructive and supportive by providing concrete examples and suggesting how to achieve improvements, not to criticise for the sake of criticising or simply as a way of demonstrating that one has completed the review task.

Practice sessions: Prior to each review task, an in-class practice session (two hours) was administered to help students experience reading and commenting on their peers' writing. Triads were formed where students exchanged their printouts and took turns to read and provide feedback on each other's work: One reviewer would read and comment out aloud to the author regarding the work, while the third student would read silently. The author was encouraged to listen, take notes, and not to defend. This form of verbal interaction was believed to be helpful in fostering writers' metacognitive strategies for self-evaluation as they listened to their peer voiced their opinions regarding their own work (cf. Tang & Thitecott, 1999). This sociocultural approach also helped students develop their thoughts during situational learning/interactions (cf. Vygotsky, 1986). During the session, the teacher served as a facilitator, presiding over

the peer feedback process and providing assistance if needed while allowing for learner autonomy.

Revision for second draft and response memo: After receiving the peers' comments, each student was asked to revise their work (second draft) based on the feedback and provide a response memo. The memo helped each student keep track of his or her actions as to what was revised, what was omitted, and what was not tackled along with a justification. The task of preparing a response memo was expected to help students reason concerning each comment and assess what would be the appropriate action to take regarding each point raised by the peer reviewers.

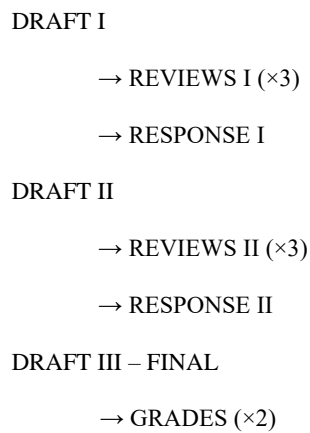


Figure 1. Organisation of the writing process involving peer feedback.

Review task two: A second-round peer review on the second draft was conducted: The same peer reviewers who read the first draft were to review the same authors' second draft. That is, the same reviewers were to review the same authors' second draft.

Revision for final version and response memo: The authors were to do a second-round revision for the final version and provide a second response memo. A process chart of the tasks was provided (see Figure 1). For practical reasons, the first draft was regarded as a practice run and thus not evaluated. The second draft and final version were evaluated.

Teacher assessors for writing performance: Two teachers served as grading assessors. Each work was assessed based on the six writing criteria (content, novelty, structure, text quality, timeliness, and references) employing a 1-5 Likert scale, 1 denoting poor and 5 excellent. In addition, the student's final draft was assigned a grade employing 1-5 Likert scale to reflect the overall writing performance.

Analysis

Wilcoxon Signed Ranks tests and Spearman rank-order correlations were performed using SPSS to analyse the scores obtained. Holm-Bonferroni adjustments were administered to correct the multi-comparison biases. Both methods are nonparametric as the dependent variables are ordinal. For ease of discussion, the second draft was

named version one and final draft as version two since the first draft was regarded as a practice run.

Table 1. Rating results for assessors A and B, version 1 and 2 in terms of mean, standard deviation (SD), rater differences (rd), version contrast differences (vcd), normalised differences (nd), Wilcoxon statistics Z, p-value, and standardised effect size r.

Criteria	Version 1			Version 2			Version contrast					
	mean	SD	rd	mean	SD	rd	vcd	nd	Z	p	r	
Content	A	2.3	1.1	-0.7	2.9	0.9	-0.38	0.5	0.8	-3.3	.007**	-.51
	B	1.6	0.7		2.5	1.0		0.9	0.5	-3.7	.004**	-.57
Novelty	A	2.3	0.8	-0.5	2.6	0.9	-0.10	0.3	0.9	-2.6	.02*	-.41
	B	1.8	0.5		2.7	0.6		0.9	0.7	-4.1	.005**	-.64
Structure	A	2.6	0.9	1.9	3.2	0.6	-0.53	0.6	1.0	-2.9	.01*	-.45
	B	1.9	0.5		2.7	0.5		0.9	0.5	3.6	.005**	-.56
Text quality	A	3.0	1.0	1.9	3.2	0.8	-0.33	0.2	0.0	-1.6	.05	-.25
	B	1.9	0.5		2.9	0.8		1.0	1.0	3.9	.006**	-.60
References	A	1.9	1.1	-0.2	2.3	0.7	0.14	0.4	0.6	-2.7	.008**	-.42
	B	1.7	1.0		2.4	1.2		0.7	0.0	-3.0	.008**	-.47
Timeliness	A	2.6	0.7	0.0	2.9	0.5	0.62	0.3	0.2	-2.4	.025*	-.38
	B	2.6	1.0		3.5	1.3		0.9	0.6	-3.8	.006**	-.59

RESULTS

Rubric Ratings

Table 1 lists the writing scores rated by the two assessors respectively and the contrasts for the scores of version 1 and version 2. Newer version yielded higher scores for all categories than the older version for both assessors. Content and references tended to be scored low in version 1 by both assessors; references remained the lowest score in version 2 by both assessors. Text quality, timeliness, and structure were among the higher scores in both versions by both assessors.

The results showed a similar trend in the six rubric scores assessed by the two assessors despite a minor variation between them. Version 2 scored mostly significantly higher than version 1 for both assessors. Assessor B's version 2 scores were significantly higher than version 1 scores in all six categories. All categories exhibited medium to large increases ranging from 0.7 to 1.0.

For assessor A, all categories gained a statistically significant score increase from version 1 to version 2 except text quality. Content and structure exhibited large increases, while novelty, references, and timeliness exhibited medium increases. Observably, the size of score increase for assessor A was smaller than that for assessor B, with the most similar increase being the references category (from 1.9 and 1.7 in version 1 to 2.3 and 2.4 in version 2 for assessors A and B, respectively). The results indicated that assessor A generally issued higher scores than assessor B, particularly for version 1, hence also contributing to smaller score increase from version 1 to version 2 compared to assessor B. References category was deemed by both evaluators to have improved to the most similar degree.

Table 2. Correlations between final grades and criteria ratings. Only significant correlations are included.

Criteria	Rater A				Rater B			
	Version 1		Version 2		Version 1		Version 2	
	<i>r_s</i>	<i>p</i>	<i>r_s</i>	<i>p</i>	<i>r_s</i>	<i>p</i>	<i>r_s</i>	<i>p</i>
Content	.59	.005**	.48	.028*	.64	.00**	.51	.017*
Novelty			.52	.015*				
Structure	.49	.024*						
Text quality	.53	.013*	.53	.014*				
References	.53	.014*	.67	.001***	.50	.02*	.59	.005**
Timeliness			.46	.035*				

Rubric Scores vs. Final Grades

Spearman rank-order correlations were also conducted to ascertain possible associations between single grades issued for the final version and scores given based on the six categories for both versions (see Table 2). The results showed that final grades when assessed by different evaluators were highly correlated ($r_s(21) = .91, p < .001$), suggesting the validity of the results. For assessor A, version 1 content, structure, text quality, and references were positively correlated with assignment grades. Assessor A's final grades and version 2 ratings were even more strongly correlated. All the version 2 categories were positively correlated with grades, except structure. These results suggest that the scores of all five categories except structure could be used to predict students' grades. The references category was the strongest grade predictor for version 2 (references > text quality > novelty > content > timeliness) and content was the strongest grade predictor for version 1 (content > text quality > references > structure). For assessor B, only content and reference showed positive correlations with grades, for both version 1 (content > references) and version 2 (references > content). Since it would be relatively quick and easy to inspect a reference list and this feature is correlated for both assessors, references may be a practical early indicator of final grades.

Table 3. Correlations between criteria ratings within the two versions for assessor A and assessor B. Only significant correlations are included.

Criteria	Rater A				Rater B			
	Version 1		Version 2		Version 1		Version 2	
	Content	Novelty	Content	Novelty	Content	Novelty	Content	Novelty
	r_s	p	r_s	p	r_s	p	r_s	p
Novelty	.63	.002 ^{***}	.81	.001 ^{***}				
Structure	.73	.001 ^{***}	.56	.009 ^{**}	.59	.005 ^{**}		
Text quality	.71	.001 ^{***}	.48	.027 [*]	.45	.041 [*]	.80	.001 ^{***}
References	.75	.001 ^{***}	.67	.001 ^{***}	.76	.001 ^{***}	.72	.001 ^{***}
Timeliness	.45	.039 [*]	.49	.025 [*]			.44	.049 [*]

Correlations among Rubric Ratings

For assessor A's version 1 scores (see Table 3), content was positively correlated with all the other categories, having a strong relationship with novelty, text quality, structure, and references and a moderate relationship with timeliness. A similar trend was also detected among assessor A's version 2 scores. Content was positively correlated with all other categories: having a very strong relationship with novelty (much stronger than

that of version 1) and moderate relationships with text quality, structure, references, and timeliness.

Considering assessor B's category scores across versions, version 1 content was strongly correlated with references and moderately with text quality. Content version 2 was strongly correlated with text quality and references and moderately correlated with structure. Novelty version 2 was moderately correlated with timeliness, and structure was moderately correlated with references and strongly correlated with text quality. Overall assessor B's scoring trend across versions was that content was at least moderately correlated with text quality and strongly correlated with references.

Table 4. Correlations between the criteria ratings of the two versions and the other assessor's final grades. Most non-significant correlations are omitted.

Criteria	Rater A criteria vs. Rater B grades				Rater B criteria vs. Rater A grades			
	Version 1		Version 2		Version 1		Version 2	
	<i>r_s</i>	<i>p</i>	<i>r_s</i>	<i>p</i>	<i>r_s</i>	<i>p</i>	<i>r_s</i>	<i>p</i>
Content	.74	.001***	.643	.002**	.57	.008**	.31	.169
Novelty	.54	.011*	.641	.002**				
Structure	.63	.002**	.48	.029*				
Text quality	.72	.001***	.70	.001***				
References	.62	.003**	.75	.001***	.32	.152	.41	.064
Timeliness			.43	.049*				

Cross-Assessor Correlations: Rubric Ratings vs. Final Grades

Cross comparisons of versions and assessors (see Table 4) showed that assessor A's final grades were moderately correlated with assessor B's version 1 content and also with a moderately yet non-significant correlation with B's version 2 references. B's version 2 content and version 1 references were comparatively the closest to A's grades, though not statistically significant. B's final grades were positively correlated with A's five categories among version 1, including strong correlations with references, structure, text quality, and content, and a moderate relationship with novelty. Assessor B's final grades were even more positively correlated with all of assessor A's version 2 category scores. These included strong correlations with novelty, content, text quality, and references, and moderate correlations with timeliness and structure. These results again suggest that content and references are the most reliable predictors of performance grades.

DISCUSSION

Do Assessment Criteria Correspond to Writing Characteristics?

Wilcoxon Signed Rank tests showed that the students' scores were significantly different in the two versions based on the six assessment rubrics of literature reviews writing, the latter version being significantly higher than the previous version, and when assessed by the two evaluators. Overall, the cross-correlation analyses indicate a highly reliable grading procedure. Despite individual preferences and differences, assessor B's grades are highly correlated to assessor A's grades and that assessor B's grades are moderately to strongly correlated to A's six rubric ratings. The findings support the validity of the assessors' evaluations and the validity of the six rubrics proposed for assessing students' writing of literature reviews. The three new rubrics introduced, i.e., novelty, timeliness, and references, appear to be helpful for guiding new research students' writing of literature reviews where newness of research is of special interest.

The novelty feature appears to be challenging for the students, as reflected among the lower scores. Assessor A gave novelty the second lowest score while assessor B rendered novelty the third lowest. Novelty requires the writing to contain new or original perspectives based on existing literature that the students have obtained. This may explain why this particular feature is given mid to low scores. It generally takes maturity to absorb information gathered and then to synthesise the new whole. As Kaufer and Geisler (1989) points out, "many of the insights required in a parameterized theory of newness have not yet made their way into theories of rhetoric or written composition". Arguably, one could assert this aspect as challenging for new research students, and it could perhaps be omitted as an assessment criterion. Still, for the learning's sake, it may be useful to contain this category. Besides, novelty improves greatly from version 1 to version 2 as observed by both assessors, and hence its educational value would benefit student writers.

Both evaluators score timeliness relatively high. Assessor A gave it the second highest of the six categories for version 1 and the third highest for version 2. Assessor B scored this as the highest for both versions. Apparently, both assessors reckoned that the specific topics that the students chose for the literature writing task were reasonable or adequate, not overly researched and still timely. This is encouraging as it is generally more engaging for the students to be able to choose their own topics from the perspectives of both peer readers and writers. One possible challenge is peer familiarity with a particular topic or subject area as a reader; however, this also applies to nearly any topic, particularly for new researchers. This category would help novice research students to carefully consider their research topics prior to their research and also during the writing process. Combined with the insight of peer readers, this dimension would also help both student writers and their peers to brainstorm other possible topics, and hence it is worthwhile to keep it as an engaging rubric.

The references category appears to be consistent when assessed by the different assessors. Assessor A gave the lowest score for both versions and assessor B gave the lowest for version 2. One may assume that it should be quite straightforward to be able to find references of sufficient quality, coverage, and recency for a given topic. The results indicated that the students seemed to struggle to find relevant and quality

references for their chosen topic, which is understandable as it usually takes time to develop good literature searching skills. Visual inspections reveal that the students used few research articles from peer-reviewed journals. In each submitted manuscript, a high percentage was based on conference papers, although some conference papers (or perhaps most) are also peer-reviewed to various degrees. Another possible reason is that library subscriptions may not contain all articles that the students initially had found, and hence they resorted to whatever sources they obtained from the Internet. It may be that information rich websites with clear yet unsupported claims are more accessible than research articles which comparatively are less accessible to the untrained eye due to the scholarly format. It should also be mentioned that the students were given insight into other strategies for acquiring papers in channels that the library was not subscribed to, including using inter-library loans and contacting the authors directly. However, these strategies may not be viewed as convenient compared to instantaneously downloading papers via the internet. Still, a plausible explanation is that the students have not developed literature searching skills. One does not obtain such skills simply by listening to one or two lectures; such skills are acquired through practice. Although finding references involves learning strategies and sometimes luck, references as a rubric may help motivate novice research students to gather relevant and important previous studies as foundations for their work.

The references and novelty features are associated with the two lower scores by assessor A for both versions. Structure and text quality are of the top two higher scores issued by assessor A in both versions. Assessor A deems students' structure and text quality with higher scores than novelty in both versions, suggesting that the novelty feature is comparatively more difficult for students to achieve. The fact that the references feature is given the lowest score in both versions implies that the students' references quality, coverage, and recentness are yet to further improve compared to the four other categories. Low scores for the references feature are consistent across assessors since assessor B also renders it the lowest and second lowest for the two versions.

Timeliness is scored the highest among the six categories in both versions by assessor B while content and references are the lowest and second lowest in either of the versions. It appears that assessor B deems timeliness the best achieved among all aspects of writing, implying that students' topics are not outdated, which is a shared view by assessor A with timeliness as the second and third highest for the two versions. Contrastively, content and references are assessed to be the lowest or second lowest in either version, hinting that these two dimensions are amongst the harder aspects to master. Indeed, both content and references require maturity for students to reflect on theories and concepts, as well as citing and interpreting relevant studies.

As found in the previous study on graduate students' social work writing (Alter & Adkins, 2006), "the evaluators agreed that students had trouble building an argumentative essay using appropriate and sufficient details drawn from the case study" based on the essay category of their six rubrics. Their result may be considered similar to that of the novelty category of the present study in that our students had difficulty establishing new or original perspectives based on the literature they gathered, particularly in version 1. In version 2, however, the novelty feature improved considerably. Alternatively, their result could be regarded as similar to that of the

content category in the present study in that students struggled to express objectives and understanding of subject areas, as the content category had the second lowest score. Nevertheless, in version 2, the content feature also improved considerably.

In the same study, the assessors “assigned relatively low scores to students' ability to write clearly focused and well-organized paragraphs” based on their paragraph rubric (Alter & Adkins, 2006). This result appears to be different from that of the text quality rubric in the present study where the students scored relatively high for both versions. One explanation may be that the students tended to use words or expressions that were quite similar to the sources they obtained, given the nature of reporting or summarising on previous studies when writing a literature review. To further clarify this possibility, a different measure should be introduced to ascertain how close the students' text is to the original sources.

Does Peer Feedback Improve Students' Writing?

The results showed that the students' final version was significantly improved compared to their first version for all the six criteria, after two rounds of peer review comments. The findings suggest that employing peer feedback as a part of the writing process does have a positive effect on students' writing of literature reviews. The findings herein thus differ from that reported by Kim (2005) where no performance improvement was found despite constructive peer feedback with comments and rational for revision suggestions. It could also be that the use of a response memo that each student is required to do in the present study has helped contribute to learners' internalisation in relation to peer feedback, where each author is required to reason and reflect over each comment received. However, Gielen et al. (2010) did not find the use of author reflection on peers' comments to be significantly contributing to learning gains. It may be that students' age, knowledge, and experience play a role in terms of reasoning maturity as their participants are secondary school students while our participants are first-year research students. In any case, peer feedback does appear to offer different insight that students could not easily obtain when working individually. Reflectively and retrospectively, to more engage students, the benefits of peer feedback should be emphasised and defended (cf. Carlson, MacDonald, Gorely, Hanrahan, & Burgess-Limerick, 2000).

Do Any of the Rubrics Predict Grade?

The correlation results showed that content, text quality, and references in both versions assessed by evaluator A are significantly correlated with grades. In the earlier less mature version (version 1), content is the most correlated with the grade, followed by text quality and references. In the more mature version (version 2), references exhibited the strongest correlation with grades, followed by text quality and content. Comparatively, novelty and timeliness have also improved across versions and changed from having no correlation with grades to being moderately correlated with grades. Contrastively, the structure category only correlated with the grades in version 1, suggesting that this could possibly be the least challenging aspect of literature review

writing. In fact, the structure could follow a template outline. Although the structure also improved from version 1 to version 2, the structure was not among the criteria that were most correlated with grades. It could be that assessor A did not view structure as the most important aspect of the literature review writing; alternatively, the importance of other categories, such as novelty, content, text quality, and references, overshadows structure. As the final grade is an overall assessment of writing performance, it would be useful to probe the assessor's mental scoring priority when evaluating writing as an aggregated score.

Assessor B's content and references are more strongly correlated with grades, suggesting that content and references are considered more important for the final grade than other categories. Content and references predict the actual grades even when assessed by the other assessor. It thus may be suggested that using references could help predict a student's performance grade, possibly more so than using content rubric as a predictor because of issues related to manuscript length. This would also help students face revisions more easily by beginning with improving references sections and citations in text that in turn also help strengthen content quality displaying depth of subject knowledge.

Implications and Future Work

This study shows that content and references tend to correlate positively with grades by both assessors. It is thus vital to enhance students' literature search skills—to find relevant studies in order to consolidate the foundations of information to be integrated, synthesised, and organised. Several sub-elements that are related to content creations could be further examined, including formations of summaries that report previous works, clarity of syntheses based on others' works, and indications of future research directions. Further work could also investigate students' provision and reception of peers' feedback as well as their internalisation process in literature reviews writing as a collaborative learning and writing process. Hence, the relationships between feedback patterns, response patterns, and performance improvement could then be explored. As student diversity and class size in higher education continue to increase, incorporating peer assessment into part of the curriculum may help enhance quality of teaching and learning, as well as institutional accountability (George-Walker & Keeffe, 2010).

Also, it would be interesting to investigate if the results herein could be applied to automatic writing scoring systems such as Automated Essay Scoring (AES) (Dikli, 2006; Attali & Burstein, 2006). Automatic feedback is a challenging yet highly interesting avenue of research; Dikli and Bleyle (2014) have already done some work on comparing automatic and manual feedback. Another interesting direction is to explore peer feedback in other aspects of language learning than writing.

CONCLUSIONS

This study explores literature reviews writing as a text genre and as a learner-centric writing process to help novice research students approach this research writing task and assist assessors (peer students or teachers) to evaluate students' writing through six

rubrics: content, structure, text quality, novelty, references, and timeliness. The writing process is accompanied by peer feedback sessions that show a positive impact on writing performance, as also confirmed in score increase. References rubric could practically serve as an early grade predictor to help motivate students to revise their work for improvement.

References

- Aguirre-Muñoz, Z., & Boscardin, C. K. (2008). Opportunity to learn and English learner achievement: Is increased content exposure beneficial? *Journal of Latinos and Education*, 7(3), 186-205.
- Alter, C., & Adkins, C. (2006). Assessing student writing proficiency in graduate schools of social work. *Journal of Social Work Education*, 42, 337-354.
- American Psychological Association. (2019). *The Publication Manual of the American Psychological Association* (7th ed.). Washington, DC: Author.
- 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.
- Attali, Y., & Burstein, J. (2006). Automated Essay Scoring with e-rater V.2. *Journal of Technology, Learning, and Assessment*, 4(3).
- 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.
- Bhatia, V. K. (2019). World Englishes in professional communication. *World Englishes*, 38(1-2), 30-40.
- Broad, B. (2003). *What we really value: Beyond rubrics in teaching and assessing writing*. University Press of Colorado.
- Bunton, D. (2005). The structure of PhD conclusion chapters. *Journal of English for Academic Purposes*, 4(3), 207-224.
- Carlson, T., MacDonald, D., Gorely, T., Hanrahan, S., & Burgess-Limerick, R. (2000). Implementing criterion-referenced assessment within a multi-disciplinary university department. *Higher Education Research & Development*, 19(1), 103-116.
- 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.
- Diederich, P. B., French, J. W., & Carlton, S. T. (1961). Factors in judgments of writing ability. *ETS Research Bulletin Series*, 2, 1-93.
- Dikli, S. (2006). An Overview of Automated Scoring of Essays. *Journal of Technology, Learning, and Assessment*, 5(1). Retrieved from <http://www.jtla.org>.
- Dikli, S., & Bleyle, S. (2014). Automated Essay Scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing*, 22, 1-17.

- 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.
- George-Walker, L.D., & Keeffe, M. (2010). Self-determined blended learning: A case study of blended learning design. *Higher Education Research & Development*, 29(1), 1-13.
- 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.
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99, 445-476.
- Hanrahan, S. J., & Isaacs, G. (2001). Assessing self- and peer-assessment: The students' views. *Higher Education Research & Development*, 20(1), 53-70.
- Hartley, J. (2003). Improving the clarity of journal abstracts in psychology. *Science Communication*, 24(3), 366-379.
- Hooper, R. I., & Butler, S. (2008). Student transfer of general education English skills to a social work diversity course: Is it happening? *Journal of the Idaho Academy of Science*, 44(2), 1-11.
- 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.
- 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.
- Hyland, K. (2002). Genre: Language, context, and literacy. *Annual Review of Applied Linguistics*, 22(1), 113-135.
- Hyland, K. (2005). *Metadiscourse: Exploring Interaction in Writing*. London and New York: Continuum.
- Karegianes, M. J., Pascarella, E. T., & Pflaum, S. W. (1980). The effects of peer editing on the writing proficiency of low-achieving tenth grade students. *The Journal of Educational Research*, 73(4), 203-207.
- Kaufer, D. S., & Geisler, C. (1989). Novelty in academic writing. *Written Communication*, 6(3), 286-311.
- Kawase, T. (2018). Rhetorical structure of the introductions of applied linguistics PhD theses. *Journal of English for Academic Purposes*, 31, 18-27.
- 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.
- 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.
- Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, 18(1), 30-43.

- Lupton, M. (2008). Evidence, argument and social responsibility: First-year students' experiences of information literacy when researching an essay. *Higher Education Research & Development*, 27(4), 399-414.
- Martin, J. R. (2009). Genre and language learning: A social semiotic perspective. *Linguistics and Education*, 20(1), 10-21.
- Mendonca, C. O., & Johnson, K. E. (1994). Peer review negotiations: Revision activities in ESL writing instruction. *TESOL quarterly*, 28(4), 745-769.
- Nygård, G. (7 January 2020). Facts about education in Norway 2020. Retrieved from <https://www.ssb.no/en/utdanning/artikler-og-publikasjoner/facts-about-education-in-norway-2020>
- Pritchard, R., J., & Honeycutt, J. (2006). Process writing. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 275-290). New York: Guilford.
- Snowball, J. D., & Mostert, M. (2013). Dancing with the devil: Formative peer assessment and academic performance. *Higher Education Research & Development*, 32(4), 646-659.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Tang, G., & Thitecott, J. (1999). Peer response in ESL writing. *TESL Canada Journal*, 16(2), 20-38.
- 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.
- Turley, E. D., & Gallagher, C. W. (2008). On the "uses" of rubrics: reframing the great rubric debate. *English Journal*, 97(4), 87-92.
- van Enk, A., & Power, K. (2017). What is a research article? Genre variability and data selection in genre research. *Journal of English for Academic Purposes*, 29, 1-11.
- Vygotsky, L. S. (1986). *Thought and language* (Revised edition). Cambridge, MA: Massachusetts Institute of Technology.
- Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: writing a literature review. *MIS Quarterly*, 26(2), xiii-xxiii.
- Xie, S. (2020). Multidimensional analysis of Master thesis abstracts: a diachronic perspective. *Scientometrics*, 123, 861-881.