Effects of anonymity on online peer review in second-language writing

van den Bos, Anne Hester; Tan, Esther

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Effects of Anonymity on Online Peer Review in Second-Language Writing

Anne Hester van den Bos, University of Applied Sciences, Leiden, the Netherlands, a.vd.bos@soverin.net

Esther Tan, Delft University of Technology, Delft, the Netherlands, e.b.k.tan@tudelft.nl

Abstract: This paper investigates the effect of anonymity in online peer review on feedback types (directive, non-directive, higher-order concern, lower-order concern) and students' revisions (processed, partly processed and not processed) in second-language writing. Participants were 114 Dutch second-year university students. They were assigned to two experimental conditions: anonymous and non-anonymous. Results showed that students in the anonymous condition provided significantly more feedback on higher-order concerns and offered significantly different types of feedback than students in the non-anonymous condition. As for revision, overall findings showed that assessees in the anonymous condition did not process more feedback (i.e., the adoption rate) than their non-identified peers, however, assessees in the anonymous condition processed significantly more directive higher-order feedback and scored significantly higher final grades for the writing module than their non-anonymous peers. These results might imply that anonymity could enable learners to provide unreservedly more higher-order concerns feedback type. On the self-same note, the adoption and revision of these higher-order feedback items was instrumental in the improved writing performance of students in the anonymous condition.

1. Introduction

1.1 Peer review as peer scaffolding

Peer review is an important aspect of collaborative learning and a critical component of a participatory culture in online learning environments. With a long history in writing instruction, peer review has been considered a crucial aspect in the improvement of writing skills (Cho & MacArthur, 2010; Yu & Lee, 2016). The peer review process requires the active involvement of students to provide feedback on each other's written text - an integrated part of collaborative and interactive learning to co-construct meaning and to negotiate shared understanding (Ajjawazi & Boud, 2017; Zhu & Carless, 2018). Unlike peer assessment where students evaluate and grade each other's writing, peer review is a formative tool where students help each other improve in areas such as logical thinking, grammar, vocabulary and other content and structural factors (Wu, Petit, & Chen, 2015). Further, providing feedback to peers is said to be more beneficial than receiving them as it is more cognitively engaging: it requires higher-order think processes to identify errors and to make suggestions for improvement (Carless & Boud, 2018). In this study, peer review and peer feedback share similar connotations, i.e., students help each other correct and improve their writing by giving feedback on each other's texts.

Peer review exemplifies Vygotsky's (1978) notion of the Zone of Proximal Development (ZPD) where a person is able to achieve a task with help from the more knowledgeable others. ZPD has its theoretical premises in Vygotsky's sociocultural theory which conceived of learning and human development as a fundamentally socially situated activity, embedded in a specific cultural environment (De Guerrero & Villamil, 2000). Recent empirical works have evidenced that not only experts can offer effective support, this form of support – scaffolding can also occur among peers (De Guerrero & Villamil, 2000; Saeed & Ghazali, 2017; Storch, 2019). In the peer review process, two or more students interact to provide each other with feedback on one another's written texts. This emergent interaction mirrors the activities in the ZPD where peer scaffolding enables an individual to accomplish a complex task which could be impossible or overwhelmingly challenging without assistance. In other words, there is a difference between what an individual can achieve alone and what the same individual can accomplish with support from someone else (Lee, 2008). Although supportive behaviour can take different forms, language is often the main tool of mediation in peer review.

Peer feedback on writing has empirically been proven to have numerous merits, particularly in "reciprocal teaching, providing peer help, receiving explanations, co-constructing ideas, resolving conflicts, and negotiating meaning" (Ge & Er, 2005, p. 146). Using scaffolding such as peer revision among second language learners has been found beneficial to learning (De Guerrero & Villamil, 2000). Several researchers have observed that second language learners can provide useful comments for successful text revision (e.g., Berg, 1999; Coté, 2014; Lu & Bol, 2007; Wu et al., 2015). Moreover, compared to expert feedback, peer feedback is often perceived to be more comprehensive since students share the same vocabulary with their peers and do not underestimate the difficulty of the task (Cho & MacArthur, 2010; Koller & Fischer, 2010; Topping, 2010). Although some research has shown that students placed more trust in expert reviews than in reviews from peers (Wu et al., 2015), Gielen et al.'s study (2010) showed that students (as opposed to experts) who participated in blinded peer review, perceived peer and expert feedback as equally helpful. The success of peer feedback might also be explained by students' knowledge

of their peers' comprehension problems as they observe their peers' learning processes from up close and often experience similar problems themselves (Cho & MacArthur, 2010; Tseng & Tsai, 2007). For instance, Ekşi's (2012) study on showed that students in the peer-review group were also able to suggest deep-level revisions on written texts as compared to students in the group that received teachers' feedback. Coté's study (2014) on peer review in an English as Foreign language (EFL) class showed that students appreciated the prospect of having their written texts reviewed by their peers and the opportunity to rectify those errors in their revised version. Students are more likely to offer feedback as well as engage with the feedback provided by learning peers (Storch, 2019).

1.2 The Peer Review Process

In the peer review process, every student assumes two roles: as an assessor who provides his/her peer with feedback on their peer's text and as an assessee who receives feedback on his/her own text. Peer review stimulates the learning process of both assessors and assessees as it increases the time spent contemplating, comparing, contrasting, and communicating about a text (Lu & Law, 2012; Van Popta et al., 2017). The act of giving feedback offers students the opportunity to engage in reflective and critical thinking (i.e., examine what constitutes a good piece of work), planning, monitoring and regulation (Rotsaert, Panadero, Schellens, & Raes, 2017). By focussing on both the form and content of the writing they might start to understand the writing process (Wu et al., 2015). Students learn effectively from reading numerous peers' work and providing feedback, while at the same time their peers' texts offer them insight into what works and what does not work for a good piece of writing (Tseng & Tsai, 2007). Assessees, on the other hand, learn by reviewing the feedback they received and by making decisions on whether to accept the feedback and how to implement changes. Research has shown that students who give more feedback and high-quality feedback are likely to be more critical of their own work (Gielen et al., 2010; Lu & Bol, 2007; Lu & Law, 2012; Lundstrom & Baker, 2009). By transferring their knowledge to their own writing, they write better texts (Lundstrom & Baker, 2009).

The peer review process is more complex than it seems. Kollar and Fischer (2010) identify three core activities in the peer review process: i) provision of feedback, ii) feedback reception, and (iii) revision. The provision of feedback is challenging because constructing useful feedback requires knowledge and skills to review, clarify, and evaluate other people's work which is a cognitively demanding task (Lu & Law, 2012; Van Gennip, Segers, & Tillema, 2010). On the self-same note, the effectiveness of peer review is contingent on the quality and the nature of the comments provided (Van der Pol et al., 2008). Carless and Boud (2018) advocate the need to develop student feedback literacy in understanding what feedback is and how to make productive use of feedback – which is instrumental in enabling student uptake of feedback. Hence, providing training and guidelines on the construction of feedback types might enhance the effectiveness of collaborative learning in peer review and students' adoption of feedback.

Next, a successful peer review process hinges on the assessee's reception of peer feedback and comments (Van der Pol et al., 2008). It is hard to predict how students would respond to peer feedback, since the willingness to follow an assessors' advice also depends on the perceived usefulness and the relevance of the feedback (Huisman et al., 2018) and trustworthiness (Topping, 2010). Moreover, before students can act on the given feedback, they will first need to fully comprehend the problems or

suggestions offered: they have to decide which comments are relevant and of high-enough quality. Vagueness has been identified as one of the reasons why feedback would not be processed (Min, 2005). Following up feedback is easier if there is sufficient explanation or suggestion (Lu & Law, 2012). Huisman et al.'s (2018) study also accentuate the importance of explanatory peer feedback which culminates in students' positive perception of peer feedback.

The final step in the peer review process is the actual revision of a text. Revision can be a demanding process as students need to compare the first draft with the given suggestions in the feedback. They have to decide either to act on those feedback items (Kollar & Fischer, 2010) or not to alter anything (Gielen et al., 2010). In other words, students have to think about the amendments they choose to make. This implies that they have to look at their text from multiple perspectives, think about possible solutions, and consider alternatives they did not notice before (Ge & Er, 2005). Research on students' actual revisions and the effect on writing quality remains limited but revision does seem to lead to improved writing performance (Cho & MacArthur, 2010; Lu & Bol, 2007).

1.3 Types of Peer Feedback and Revisions

Feedback in peer review differs in extent, content and style. Various researchers have used different coding criteria to categorise different types of feedback. Some researchers distinguish between cognitive feedback that targets the content of the work and affective feedback which points out the quality of the work in terms of general praise or criticism (Nelson & Schunn, 2009). Other researchers made a distinction between feedback provided on a local and on a global level (e.g., Liou & Peng, 2009; Lundstrom & Baker, 2009; Min, 2005). Van Steendam, Rijlaarsdam, Sercu, and Van den Bergh (2010) used a similar categorisation identifying feedback on higher-order level and lower-order level. They explained that global or higher-order concerns (HOC) affect larger portions of the text such as idea development, audience and purpose, style, organisation and argumentation while local issues or lower-order concerns (LOC) target wording, grammar, spelling or punctuation. Min (2005) concluded that second-language learners need feedback on both local and global level in order to improve their writing skills. This is because, unlike first-language learners, they struggle to organise and express their ideas in English (higher-order) and compose rich sentences (lower-order). Min's study (2005) also showed that without training, second-language learners often focussed on local issues such as spelling and grammar rather than global issues such as structure and content. This was true for provision of feedback as well as for revision. Van Steendam et al. (2010) arrived at similar conclusions that rendering feedback on higher-order concerns was not easy and without training most second-language assessors would focus on the surface level of a peer's text.

Other scholars such as Cho, Schunn and Charney (2006) and Cho and MacArthur (2010) distinguish between directive feedback that contains explicit suggestions or specific changes in a student's written text and non-directive feedback that involves more general, nonspecific observations that could apply to any text. Cho et al. (2006) claimed that directive comments might lead to changes in the text, but not to changes in a student's writing behaviour. A possible explanation is that students who receive directive feedback will simply accept the given suggestions whereas students who receive non-directive feedback have to think about the possible revisions and solve the problems themselves, thus developing their self-reliance.

Instead of categorizing feedback types, some researchers chose to focus on the types of revisions students made in the review process. Researchers have used various criteria for categorizing revisions: focussing either on the actual revisions (Paulus, 1999; Min, 2006; Van der Pol et al., 2008), on a holistic scoring on the quality of the final drafts (Gielen et al, 2010) or on both (Cho & MacArthur, 2010). Revision can be measured by focussing either on the quantity or the quality of the feedback and revisions. The first option is to focus on the willingness to incorporate the given feedback by counting the number of peer comments that are accepted and used for revision, i.e., the adoption rate. Several researchers have used this adoption rate but results have been contradictory. Some noted a large percentage of incorporated feedback (Coté, 2014; Min, 2006) whereas others found a much smaller percentage (Liou & Peng, 2009; Paulus, 1999). Moreover, there is no evidence that the number of revisions correspond to the quality of revision (Paulus, 1999; Min, 2006).

The second option is to focus on the quality of the feedback and revisions. For example, Liou and Peng's (2009) study showed that not all feedback culminated in improved writing and revision quality was associated with the effectiveness of revisions rather than with the adoption rate. On revision quality, Cho and MacArthur's (2010) study showed that complex repairs were associated with non-directive feedback which led to improved writing performance (partial $\eta_2 = 46$), whereas directive feedback often led to simple repair revision. Their study showed that the types of feedback have a significant bearing on revision quality.

A third option is to consider student's text revisions based on global and local comments. Saeed and Ghazali's (2017) research on asynchronous and synchronous online group review found that local comments facilitated local text revisions while global comments led to global text revisions. In addition, they labeled students' text revisions either as 'comment-based' or 'self-made' and concluded that global text revisions were more likely to be connected to peers' comments (46%), than local text revisions (10%). In other words, global comments facilitated students' global text revisions and helped them to refine their texts whereas for local text revision to occur, peer comments were less essential. This supports the idea that second language learners should be stimulated to provide feedback on global, higher-order level rather than on lower order concerns (Min, 2006; Van Steendam et al., 2010).

1.4 Anonymous and Non-Anonymous Peer Review

As aforementioned, peer review is an integral part of a collaborative learning process (Van der Pol et al., 2008; Van Gennip et al., 2010). Peer feedback is a core activity in the peer review process: the act of providing peer feedback is conceived of as a collaborative activity where peers feel a sense of mutual dependence to complete a shared task. Such a student-centred learning experience requires psychological safety, value diversity, and interdependence (Van Gennip et al., 2010). When students trust their peers and believe that the environment is safe for interpersonal risks, they are more likely to regard differences of opinion as a starting point for learning (Topping, 2010). In classroom settings, non-anonymous peer review, involving students who know one another's identity, has been standard practice. However, anonymous peer review, in which both assessors and assessees are unknown to each other, could mitigate the influence of status differences, friendship or retribution (Cho & MacArthur, 2010). For assessors, not knowing the assessees' identity can alleviate some of the uneasiness caused by social pressure and the reluctance to criticize others as found in cooperation-oriented cultures (Liou & Peng, 2009). Accordingly, students

in higher education indicated that they felt more comfortable giving feedback anonymously in peer assessment (Raes et al., 2015). In a quasi-experimental study by Vanderhoven et al. (2015) on peer assessment in secondary education, they found that students in the anonymous group – in which assessors assessed their peers in class without revealing their identity, using an electronic voting system - experienced significantly less peer pressure and less fear of disapproval as compared to students in the identifiable group. On a similar note, Lu and Bol's (2007) study showed that the use of anonymity in e-peer review process, in which both reviewers' and reviewees' identities were unknown to each other, resulted in better students' writing performance in the post-test scores ($\eta_2 = 0.19$). Students seem to be more honest and critical, possibly because they feel safer in expressing their opinions, as they do not need to worry about the author's feelings. Furthermore, Liou and Peng's (2009) study on computer-mediated peer review, found that students looked forward to the prospect of writing for an unknown audience. Anonymity also affected the drafting process as students planned more extensively and wrote more carefully when they were communicating with an audience of unknown peers than when they were evaluated solely by instructors (Lu & Bol, 2007).

Anonymous peer feedback in the revision phase can address the issue of students mistrusting weaker students' comments (Paulus, 1999). Many students hesitate to accept peers' feedback when they know their peers are less capable writers than themselves, even if the comments are correct (Lu & Bol, 2007). One of the aspects that affects students' uptake of feedback is the trust in their peers' ability as assessor. Van Gennip et al.'s (2010) study found that peer assessment intervention led to significantly higher scores on trust (Cohen's $d = .44$): the higher degree of trust students had on their assessors, the more positive their perceptions of peer assessment became. Liou and Peng's empirical work (2009) evidenced that students who did not trust the effectiveness of peer review but who adopted the feedback anyway, did successfully revise their writings and ended up submitting texts of better quality. An anonymous distribution might provide a sufficient degree of uncertainty regarding the peer's status, age, past grades, gender and language proficiency (Johnson, 2001). It might also induce a mindful and critical acceptance of the received peer comments that asks for more deep thinking before accepting or rejecting the feedback (Gielen et al., 2010; Huisman, Saab, Van Driel, Van den Broek, 2017). In a study on college writing in the first Language (English), Lu and Bol (2007) found notably higher post-test scores on writing performance in the anonymous condition than in the non-anonymous condition. Similar findings were established by Guilford (2001) whose students claimed that the quality of their term papers and their course grades improved after using anonymous peer review approach during the writing process.

Research has shown that the perception of feedback is important (Strijbos et al., 2010). However, the intricate relationship between types of feedback and the way feedback is received and processed remains unclear (Van der Pol et al., 2008). Furthermore, the impact of anonymous and non-anonymous conditions on feedback types has hardly been studied. Likewise, there is paucity of research on the impact of the various types of feedback on students' actual revisions and the effect on writing quality. Besides, with the advent of computer mediated technology, classroom practices have become increasingly blended and this has created the possibility of students offering each other online peer review. This has also enabled students from different classes to collaborate online synchronously and asynchronously. Although positive effects have been found for the use of computer-mediated technology in the peer review process (Liou & Peng, 2009), empirical research on the impact of

instructional interventions, conditions and circumstances of online peer review on effectiveness and learning remains limited (Cho & MacArthur, 2010; Van Popta, Kral, Camp, & Martens, 2017).

Notwithstanding the rich potential of peer review, a main drawback of the peer review process is the undesirable social effects that are inherent to the process. The interpersonal context that comes with peer review leaves room for peer pressure due to friendship bonds or enmity, and fear of disapproval (Topping, 2010; Vanderhoven, Raes, Montrieux, Rotsaert, & Schellens, 2015). Several studies found that students felt uncomfortable criticising each other's work (Liou & Peng, 2009; Raes, Vanderhoven, & Schellens, 2015; Topping, 2010) while others indicated that students found it extremely difficult to give negative feedback to classmates, especially friends, because they were concerned about hurting others' feelings or damaging personal relationships (Lu & Bol, 2007). Likewise, literature provides few empirical studies on the quality criteria of received feedback, students' revision and performance improvement (Strijbos, Narciss, & Dünnebier, 2010). It is unclear what types of learning settings and instructional scaffolds in an online learning environment would foster effective and successful peer feedback in second language writing. In essence, we were interested to find out if the identity of assessors and assesses were kept confidential - could anonymity be instrumental in the provision of feedback types with a special focus on higher-order feedback types as these have often been considered to be more critical and useful in the peer review process. Likewise, we wanted to know if anonymity would affect the successful uptake of feedback in the revision process. This empirical study aims to investigate the effect of anonymity on online peer feedback, feedback types and students' revision of their written texts in English as Second Language (ESL) writing.

2. Research Questions and Hypotheses

Building on previous research as afore-discussed, this study investigates the effect of the anonymous and non-anonymous online peer review process on Dutch university students in an English as Second Language writing class. We examined the effect of anonymity on: i) the different types of feedback (directive (D) vs. non-directive (ND) and higher-order concerns (HOC) vs. lower-order concerns (LOC) as well as a combination of D_HOC, D_LOC, ND_HOC & ND_LOC feedback types; ii) students' adoption of feedback in the revision of written work; and iii) students' writing performance. Following are the research questions and hypotheses:

RQ1: What is the effect of anonymous and non-anonymous online peer review on the different types of feedback?

> H0: There is no significant difference between assessors in the anonymous and non-anonymous conditions on the provision (in terms of quantity) of directive and non-directive higher-order concerns (HOC).
>
> H1: Assessors in the anonymous condition suggest more critical and substantial changes. They provide more feedback on higher-order concerns (HOC) than assessors in the non-anonymous condition be it directive or non-directive higher-order concerns.
>
> H0: There is no significant difference between assessors in the anonymous condition and non-anonymous condition on the provision (in terms of quantity) of the four different feedback types (D_HOC, D_LOC, ND_HOC & ND_LOC).

H2: Assessors in the anonymous condition provide more feedback items in the four different feedback types (D_HOC, D_LOC, ND_HOC & ND_LOC) than assessors in the non-anonymous condition.

RQ2: What is the effect of the anonymous and non-anonymous online peer review and feedback types on students' revisions?

H0: There is no significant difference between the anonymous and non-anonymous conditions on the number of feedback items processed by students.

H3: Students in the anonymous condition process more feedback items than their peers in the non-anonymous condition, i.e., they show a higher adoption rate.

H0: There is no significant difference between the number of directive lower-order concerns (D_LOC) processed and the number of non-directive higher-order concerns (ND_HOC) feedback.

H4: Feedback on directive lower-order concerns (D_LOC) is more likely to be processed than non-directive higher-order concerns (ND_HOC) feedback.

H0: There is no significant difference between the anonymous and non-anonymous conditions on the number of directive and non-directive higher-order concerns (D_HOC, ND_HOC) processed by students.

H5: Assessees in the anonymous condition process more feedback items of directive and non-directive higher-order concerns (D_HOC, ND_HOC) than students in the non-anonymous condition.

RQ3: What is the effect of anonymous and non-anonymous online peer feedback and revisions on writing performance?

H0: There is no significant difference between the anonymous and non-anonymous conditions on students' final grades in the writing performance.

H6: Students in the anonymous condition score higher final grades in writing performance than students in the non-anonymous condition.

3. Method

3.1 Participants and Setting

All 126 second-year university students of a University of Applied Sciences in the Netherlands, age 18 to 24 years (average age 21) were invited to participate in this quasi-experimental study. However, owing to absentees and dropouts, the eventual number of participants was 114. All students were native speakers of Dutch, there were no bilingual students with English as a mother tongue. All students had acquired B2 level in English of the Common European Framework of References for Languages which indicates that all students had reached an upper intermediate level in reading, listening, speaking and writing. Table 1 provides an overview of the number of assessors and assesses in the two experimental conditions. Most of the students, not all, received feedback from two peers. 114 students offered feedback to peers (assessors) while 100 students revised their essays based on the feedback they received (assessees). At the start of the academic year, students were randomly put into five classes by the university administration. Subsequently, each class was split into two groups by the administration (group A and B) for a number of different courses, including the English writing module in which the research took place. All A groups were assigned to the non-anonymous setting, and all B groups made up the anonymous condition. Assessees and assessors in the non-anonymous setting knew one another's identity whereas the identity of the assessors and assessees were

unknown to each other in the anonymous setting. However, in general, most of the students in both experimental conditions knew each other as they had been working together intensively in a number of courses since the start of the academic year. In line with the protocol from the Research Ethics Committee (cETO), informed consent was obtained from all participants.

Table 1.

Assessors and assessees in the two experimental conditions

	Assessors			Assessees		
	Anonymous condition	Non-anonymous condition	Total	Anonymous condition	Non-anonymous Condition	Total
Male	41	38	79	32	35	67
Female	22	13	35	21	12	33
	N=63	N=51	N=114	N=53	N=47	N=100

3.2 Materials

All students in the anonymous and non-anonymous condition followed an 8-week course called English Writing Power 2. Lesson procedure, materials and teacher facilitation were similar for both conditions. The free online programme used in this course was called Peergrade. It offered the possibility for all students to follow exactly the same online procedures. The only difference was that students in the anonymous condition saw "submission 1" and 'submission 2", whereas their peers in the non-anonymous condition saw their peers' names (see Figure 1).
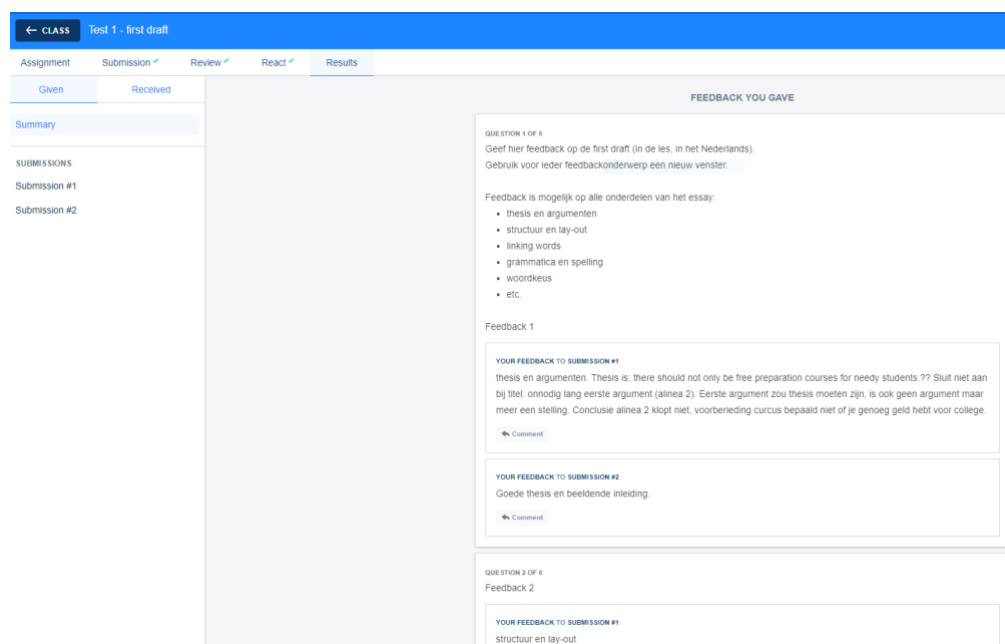

Figure 1. Screen shot of students' peer feedback in the anonymous condition.

Students could provide feedback and upload the first draft and the final draft of the essays. All feedback items were categorised according to the different types of feedback and checked for revision.

The 8-week course known as Writing Power 2 started in the first week of February 2018 (See table 2 for an overview of the lesson phases). All students followed the same weekly one-hour face-to-face classes, the same instructional programme and in-class activities. Attendance was mandatory for seven out of eight classes. In the first four weeks, students were trained in writing, argumentation and peer review. The benefits of peer review were explained at the beginning of the course. The differences between directive and non-directive feedback and higher-order and lower-order feedback were also exemplified. To practise writing and peer feedback, students wrote short texts and a group essay on which they provided peer feedback in class. From week five to nine, students had to write two five-paragraph argumentative essays of circa 500 words on ethical business cases where both essays were part of the test. Essay 1 accounted for 40% of the final grade (20% persuasion and coherence of the essay, 20% peer feedback and revision). Essay 2 made up the other 60% (persuasion, coherence, grammar, spelling, word use) of the final grade. Students' writing performance was derived from these final grades for Writing Power 2. Final grades were given on a scale of one to ten: 1 = very poor, 10 = excellent. The exam and assessment rubrics were similar for all students. Only the first essay - Essay 1 was used for this research study.

The assignment for essay 1 was to write a five-paragraph essay on an ethical business case which students could choose from the website - cases.ethicsworkshop.org. Students received the details of the assignment in week 4; discussed it in class and after a short demonstration of the online programme Peergrade, they uploaded a framework of their text to Peergrade. In week 6 they gave each other online peer feedback on the frameworks (the basic structure of the essay, i.e. thesis statement and three arguments) of essay 1 in class. Using the feedback they received on the framework, students wrote the first draft of their essay (five paragraphs; circa 500 words) and uploaded their text to Peergrade at the start of week 7. In week 7 students participated in the actual peer review process by providing online peer feedback to two peers in class. The teacher was available for practical questions on the assignment and on the use of the programme but did not intervene in the peer review process. The majority of students gave peer feedback on two texts, and consequently most students received feedback from two fellow students. After class, students received a notification from the programme that they could view the given feedback online and use it for revision. In week 8 all students uploaded their final drafts of the first essay. Both essay 1 and essay 2 with a signature and a declaration of originality were handed in and uploaded to a programme that checks for plagiarism. Instruction as well as students' feedback comments were in Dutch so that language would not become a barrier or a confounding variable in the study. Besides, Lee (2008) regarded the use of native language as a mediating tool for cognitively demanding tasks to be appropriate as it reduced the cognitive burden. Two teachers facilitated the Writing Power course. Both teachers had a Master's degree in English language and culture, similar work experience and both had taught the course Writing Power in previous years. One teacher taught four groups (two anonymous, two non-anonymous), the other taught six groups (three anonymous, three non-anonymous). The latter also conducted the research.

Table 2.
Overview of lesson phases in the two experimental conditions

Lesson Phases	Anonymous Setting	Non-anonymous Setting
Week 1 – 4: classes on writing, (ethical) argumentation and training in peer review. Students wrote a group essays. (Week 5: no class)	4 x 1 hour	4 x 1 hour
Week 6: In class: students gave online peer feedback on frameworks essay 1	1 hour	1 hour
Week 7: In class: students gave online feedback on first drafts essay 1	1 hour	1 hour
Week 8: students scored the feedback they received + revision. Students uploaded the final draft of essay 1	1 hour	1 hour
Week 9: Students wrote essay 2 (take-home assignment)	8 hours	8 hours
Total Duration	7 hours class 8 hours test	7 hours class 8 hours test

3.4 Measures and Analysis

The entire corpus of data was analysed using Chi's (1997) verbal analysis framework to determine a unit of analysis according to its semantic features. For example, if students offered feedback containing a number of different topics, these were segmented into singular units for analysis based on one topic/ idea. An example of a cluster of feedback statements that was segmented into singular units is presented in table 3. In total, 1490 units of analysis were identified: 835 from the anonymous condition and 655 from the non-anonymous condition.

Table 3.
Example of a string of feedback segmented into single units for analysis

Example in English (translated)	Original example: classifying feedback items
Paragraph 2: Your First argument is good only I will formulate 'can give it' differently // I would place a comma for especially because it refers to the previous sentence // Coca Cola has to be written with capitals;) // In the last sentence I would not use the word demage because their way of looking at Coca Cola is not damaged but changed negatively.	"Alinea 2: Je eerste argument is goed alleen ik zal zelf 'can give it' anders formuleren // Ik zou voor especially een komma zetten aangezien die zin over de zin daar voor gaat. // Coca Cola moet met hoofdletters;)// Ik zou in de laatste zin niet het woord demage gebruiken aangezien hun manier van kijken naar Coca Cola niet wordt beschadigd maar wel negatief veranderd."

Next, all the single feedback items were filtered into three categories. The first category contains feedback that indicated a possible change (Action) or correction to

improve the text, in other words, the revision-oriented comments. According to Parr and Timperley (2010), only feedback items that suggest changes could be followed up. Although feedback in the form of praise, criticism or summary has in some cases been found to help a student's writing process (Cho et al., 2006), they have rarely proven to have an effect on writing improvement in other research studies (Cho & MacArthur, 2010). In total, there were 275 such feedback items that were not used for further analyses. The second category was feedback items that contained repetition of comments. These feedback items (27 of 1490) were labelled (H) and were not used for analyses. The third category was feedback that indicated a missing title or a complete paragraph owing to incomplete work from the assessees. These feedback items were labelled (R) and not used for analyses (87 of 1490 feedback items).

Finally, the remaining 1103 feedback items were first categorised into four different types of feedback: higher-order concerns (HOC), lower order concerns (LOC), directive (D) and non-directive (ND) (see Table 4). Lower-order concerns involved spelling, grammar, targeting at wording and punctuation. Higher-order concerns included development of ideas, audience and purpose, style, organisation and argumentation (Liou & Peng, 2009; Van Steendam et al., 2010). Directive feedback contained explicit suggestions or specific changes in a student's written text (Cho & MacArthur, 2010). In other words, directive feedback items offered a problem plus a solution. Non-directive feedback on the other hand included observations and comments without suggesting any specific changes. Next, these categories were combined resulting in four feedback types, i.e., directive feedback on lower-order concerns (D_LOC), non-directive feedback on lower-order concerns (ND_LOC), directive feedback on higher order concerns (D_HOC), and non-directive feedback on higher order concerns (ND_HOC) (see Table 4 for a detailed description of the four categories of feedback types). A second rater was asked to code the feedback items for action (FBAction) and the different types of feedback (FBTypes). Cohen's K was run on a sample of 12.2% of the feedback items to determine if there was inter-rater agreement. Cohen's Kappa for revision-oriented feedback (FBAction) was κ = .98 (12.2%), p = .00 and for the four types of feedback (FBType) κ= .78 (9.1%) p = .00 which indicated a substantially high inter-rater reliability.

Table 4.
Overview of the different feedback types and sample feedback items

Categories		Directive	Non-directive
	Description	D: Explicit suggestions or specific changes in a students' written text	ND: Nonspecific observations, including all comments on details without suggesting a specific adjustment.
Higher-order concerns	HOC: Development of ideas, audience and purpose, style, organisation and argumentation	Example D_HOC: "For more structure in the text, I would use firstly, secondly... etc." *(Translation: "Voor de structuur in je tekst zou ik argumenten inzetten met firstly, secondly...etc.)*	Example ND_HOC: "The second argument does not agree with the thesis statement." *(Translation: "Het tweede argument sluit niet aan op je thesis*

			statement,")
Lower-order concerns	LOC: Spelling, grammar, targets wording, lay-out and punctuation	Example D_LOC: "Besides, it sounds better if you use Persons instead of People. (Translation:"Daarnast klinkt het beter als je Persons vervangt door People.")	Example ND_LOC: "But then is informal language so change that. (Translation: "But then is een spreektaal, dus pas dat aan.")

Revisions in students' texts were measured on an ordinal scale of three categories: processed, partly processed and non-processed (See table 5 for a more detailed description), which is a simplified version of Cho and MacArthur's scheme (2010) on simplicity and complexity of revision. Students' first drafts were compared to their final drafts using 'compare documents' in Word. Each of the 1103 feedback comments was analysed to determine if revision was made as well as the extent of the revision: more than half of the feedback items were processed (55.8%), slightly over 10 percent (10.2%) was partially processed and about one third was not processed (34%).

Table 5.
Coding categories and description of the quality of revisions

Categories	Descriptor
Not processed	The final draft did not show any sign of change that could be linked to the specific peer feedback comment Example: First draft "…the government must be harder to let marketing go on in a fair way." Feedback: "the government must be harder" is Dutch translated into English *(Translation: "the government must be harder" is Nederlands vertaald naar het Engels)"* Final Draft: "The government must be harder…"
Partly processed	The final draft showed some signs of changes, however, the problem indicated by the specific feedback item was not solved Example: First draft (refers to the third argument in a text) "There should come a clear line between the ethically acceptable civil disorder and going too far during strikes." Feedback: The third argument is a new thesis statement, it is not an argument that supports your thesis statement. *(Translation:"Het derde argument is een nieuwe thesis statement (stelling), het is geen argument voor je thesis statement")* Final Draft: "Thirdly, how should local police departments respond to civil disobedience?"
Processed	The final draft showed clear signs of a change related to the specific peer feedback item Example: First draft (refers to the first argument in a text)" Explicit rules can prevent firing." Feedback: "The first argument can be clearer, what do you mean by 'rules'?" *(Translation: ,"Eerste argument kan duidelijker, wat bedoel je met regels?)*

To answer the research questions, three main statistical analyses were conducted. Independent samples t-tests were first used to compare the means of continuous variables between the anonymous and non-anonymous conditions. Analysis of the relationships between the categorical variables: the two experimental conditions (anonymous / non-anonymous), types of feedback (D_LOC, ND_LOC, D_HOC, ND_HOC) and revision (not processed, partly processed, processed) were performed using Chi-square tests. In addition to the chi-square analyses, z-tests were run to compare column proportions of the four feedback types in the two experimental conditions. And multinomial logistic regressions were performed to determine if prediction of the nominal dependent variable 'revision' was possible based on independent variables: type of feedback and the two experimental conditions.

4. Results

Overall, 1103 revision-oriented feedback items were coded: 620 anonymous and 483 non-anonymous. No statistically significant differences were found in the number of revision-oriented feedback items (FBAction) provided by students in the anonymous condition (M= 6.08, SD=3.74) and in the non-anonymous condition (M=5.55, SD=3.43), t(187) = -1.00, p = .318 (two-tailed). The succeeding sections address each of the three research questions.

4.1 Anonymity and Feedback Types

The first research question looked at the effect of anonymous and non-anonymous online peer review on the higher-order concerns (including both directive or non-directive higher-order concerns). The first hypothesis claimed that students in the anonymous setting would offer more critical and substantial feedback, i.e., higher-order concerns, than students in the non-anonymous setting. On average, the students participating in the study offered each other just over three feedback items on higher-order concerns (M = 3.10, SD = 1.90). Independent sample t-test was run to compare the means of the number of feedback items on higher-order concerns between the anonymous and the non-anonymous condition. Findings showed that the students in the anonymous condition provided significantly more feedback items on higher-order concerns (M = 3.37, SD = 1.76) than students in the non-anonymous condition (M = 2.79, SD = 2.02), t(164) = -1.97, p = .026, Cohen's d = .31 (one-tailed) Hence, the null hypothesis was rejected.

Hypothesis two predicted that students in the anonymous condition would provide their peers with more number of feedback items in all the four types of feedback (D_LOC, D_HOC, ND_LOC, ND_HOC) than their non-anonymous peers. Figure 2 provides an overview of the frequency of occurrences of the four different feedback types.
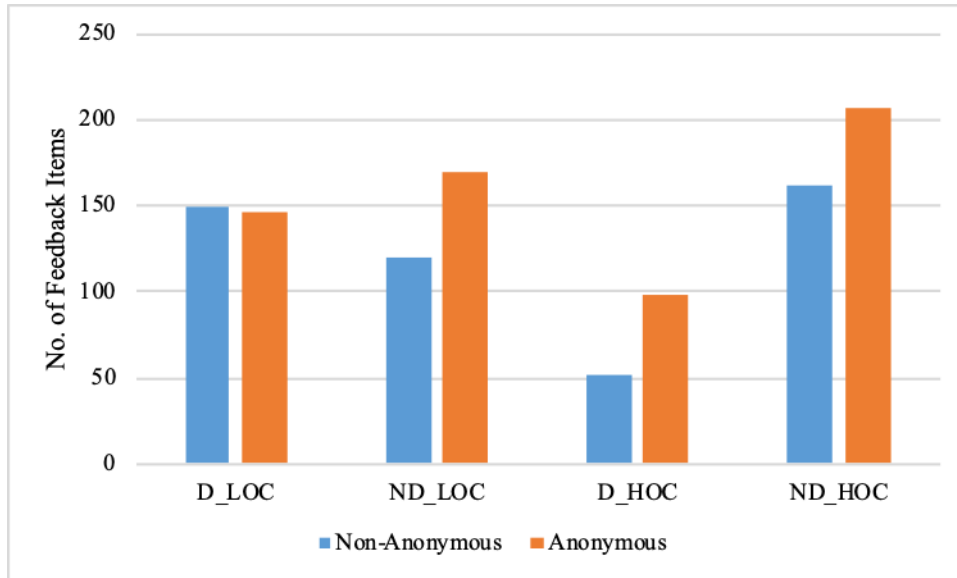
Figure 2. Frequency of occurrences of feedback types in the two experimental conditions

Note: D_LOC = Directive lower-order concerns, ND_LOC = Non-directive lower-order concerns, D_HOC = Directive higher-order concerns, ND_HOC = Non-directive higher-order concerns

Chi-square analyses was first performed to determine whether there was a relationship between the two experimental conditions and the four types of feedback. Analysis indicated that there was a significant correlation between the anonymous and non-anonymous and the four different feedback types, χ_2 (3, N = 1103) = 11.93, p = .01, V = .104. Next, a z-test was run to compare the column proportions of the four feedback types in the two experimental conditions (see Table 6). Findings showed that the proportion of directive lower-order concerns (D_LOC) was significantly lower in the anonymous condition than in the non-anonymous condition whereas, the proportion of directive higher-order concerns (D_HOC) was significantly higher in the anonymous condition than the non-anonymous condition. There was no statistical significance for the other two feedback types in both experimental conditions. Hence, hypothesis two was partially accepted.

Table 6.

Comparison of column proportions of the four feedback types in the anonymous and non-anonymous condition

Feedback Types	Anonymous	Non-Anonymous
	% within condition	% within condition
D_LOC	23,5%b	31,1%a
ND_LOC	27,4%a	24,8%a
D_HOC	15,8%b	10,6%a
ND_HOC	33,2%a	33,5%a
Total	100%	100%

Note: D_LOC = Directive lower-order concerns, ND_LOC = Non-directive lower-order concerns, D_HOC = Directive higher-order concerns, ND_HOC = Non-directive higher-order concerns

15

The second research question focussed on the effect of anonymity and feedback types on revision (i.e., processed, partly processed, not processed). Hypothesis three stated that assessees in the anonymous condition would be more likely to process feedback items than their peers in the non-anonymous condition. In other words, the adoption rate would be higher for anonymous assessees. However, chi-square test indicated that there was no statistically significant difference between the experimental conditions and revision types χ_2 (2, N = 1103) = 0.70, p = .71. The null hypothesis was accepted as assessees in the anonymous condition did not process more feedback items than students in the non-anonymous condition.

The fourth hypothesis predicted that directive lower-order feedback (D_LOC) would be easier to process, and therefore, more likely to be used for revision than non-directive higher-order feedback (ND_HOC) which is considered more complex. Figure 3 provides an overview on the extent of the revision (not processed, partly processed & processed) in the four feedback types. Descriptive statistics showed that more directive lower-order concerns (D_LOC) were processed as compared to non-directive higher-order concerns (ND_HOC). Chi-square indicated a significant association between the four feedback types and revision χ_2 (6, N = 1103) = 118.97, p = .00. Next, multinomial logistic regression was run to investigate the relationship between the outcome variable, i.e., the revision based on the feedback items and the predictor variables – the four different feedback types. The results showed a significant chi-square χ_2 (6, N = 1103) = 124.87, p = .00. The parameter estimates for feedback that was processed indicated that the probability of directive, lower-order feedback (D_LOC) feedback being processed compared to not processed was higher than non-directive, higher-order concerns (ND_HOC), b = 1,412, Wald χ_2 (1, N = 1103) = 57.99, p = .00. Consequently, the fourth hypothesis was accepted.
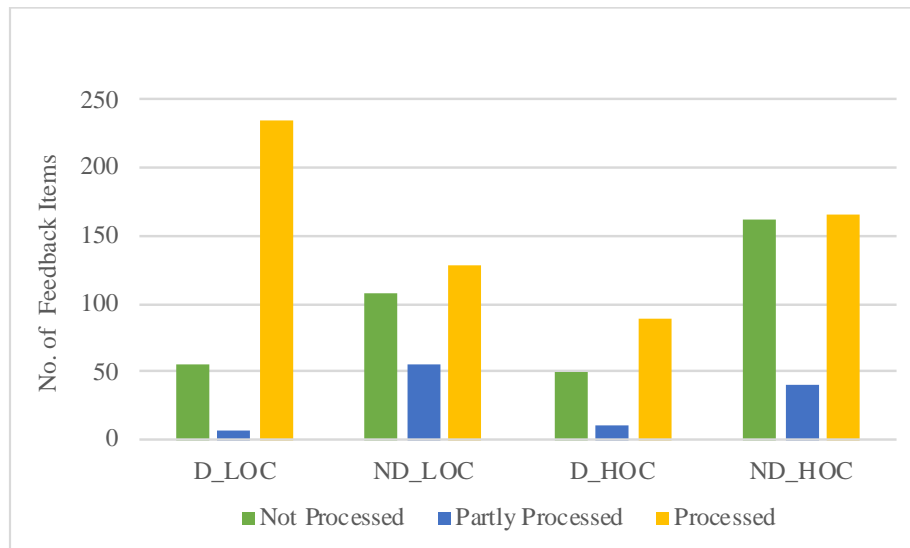

Figure 3. The extent of revision in the four feedback types.
Note: D_LOC = Directive lower-order concerns, ND_LOC = Non-directive lower-order concerns, D_HOC = Directive higher-order concerns, ND_HOC = Non-directive higher-order concerns

Hypothesis five stated that students in the anonymous condition would process more feedback items of higher-order concerns (D_HOC, ND_HOC) than students in the non-anonymous condition. Figure 4 provides an overview of the number of

feedback items that were not processed, partially processed and processed in the two experimental conditions. Multinomial logistic regression showed that the four different feedback types and experimental conditions (anonymous/non-anonymous) explained a significant amount of the variance in students' revisions, χ^2 (14, N = 1103) = 137.88, p = .00. Students in the anonymous condition significantly processed a larger quantity of feedback items on directive higher-order concerns (D_HOC), b = .551 Wald χ^2 (1) = 4.33, p = .04 as compared to the non-anonymous condition. There was no statistical significance for non-directive higher-order concern (ND_LOC). Anonymity positively predicted the adoption and the revision of directive higher-order concern feedback (D_HOC). Hypothesis five was partially accepted.
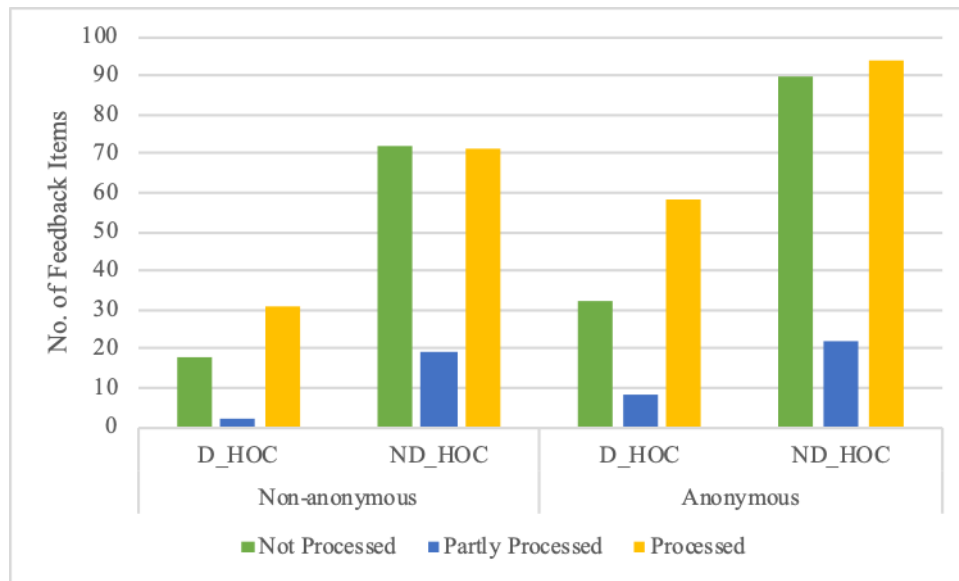

Figure 4. The measure of revision of HOC (higher-order concerns) feedback type in the two experimental conditions.

Note: D_LOC = Directive lower-order concerns, ND_LOC = Non-directive lower-order concerns, D_HOC = Directive higher-order concerns, ND_HOC = Non-directive higher-order concerns

4.3 Anonymity and Writing Performance

The third research question addressed the effect of anonymous and non-anonymous onine peer feedback on writing performance. Hypothesis six predicted that students in the anonymous condition would score higher grades than students in the non-anonymous condition. An independent t-test showed that students in the anonymous condition received significantly higher grades for the writing module (M=7.43, SD=1.12) than students in the non-anonymous condition (M=6.94, SD=1.03), t(187) = -3.099, p = .001, cohen's d = .46 (one-tailed). Hence, the null hypothesis was rejected.

4.4 Summary of the main findings

In sum, we found that there were statistically significant positive relations between anonymity and online peer review. The three critical findings were: first, the anonymous assessors provided significantly more feedback on higher-order concerns such as development of an idea and structuring an essay than their non-anonymous peers. Second, directive, higher-order feedback (D_HOC) was significantly more processed by assessees in the anonymous condition than those in the non-anonymous

17

condition. Third, students in the anonymous condition scored higher final grades than students in the non-anonymous condition.

5. Discussion

This study investigated the effect of anonymity on the online peer review process. It was hypothesized that anonymous and non-anonymous online peer review would render different effects on the types of feedback and revisions, and consequently, students' writing performance. The 114 Dutch second-year university students who participated in this research study provided their peers with online peer feedback on the English essays. This section discusses the critical theoretical and pedagogical implications underlying these results of this empirical work.

Essentially, the present study investigated three related research foci: provision, adoption and revision of higher-order concerns feedback types in both anonymous and non-anonymous condition. Findings showed that anonymity did indeed yield more occurrences of feedback items in both directive and non-directive higher order concerns. However, it is also noteworthy that prior to the research implementation, students in both experimental conditions received peer feedback training which created awareness of what good feedback was and how to render good feedback (Rollinson, 2005; Van der Pol et al., 2008). Students were trained over a period of four weeks in writing and in developing argumentation in the peer review process: differences between directive and non-directive feedback and higher-order and lower-order feedback were highlighted in the training phase. Next, we attributed the provision of more higher-order feedback items to the lack of inhibition as anonymous assessors were giving feedback to an unknown audience. One critical element of the ZPD is the psychological developmental progress - it posits that an induction first occurs on an interpsychological level and only second on an intrapsychological level, i.e., "a child's experiences of participating in activities are first externally accessible that the structures and content of mental life that can be played out internally" (Pea, 2004, p. 426). Here, modes of individual thoughts have to be internalized from communicative interactions with other people (Stahl, 2005) and individual learning occurs where one internalizes or externalizes knowledge that was first constructed with others. Thus, anonymity in peer review could have deterred resistance to a first internalization of knowledge before an externalization of knowledge could be expected to occur. Anonymous assessors were able to provide explicit suggestions / comments for change as they were less restrained in exercising more in-depth and critical thinking in the construction of feedback.

On the adoption and the revision of feedback types, findings showed that overall anonymous assessees did not process more feedback than their non-identified peers and there was no statistical significance in the adoption rate for both experimental conditions. However, both provision and processing of feedback on higher-order concerns was affected by anonymity. This seems promising since different researchers have shown interest in how to stimulate second-language students to provide feedback on global, higher-order level rather than on the 'easier' local or lower-order concerns (Min, 2006; Van Steendam et al., 2010). On this note, it is also important to highlight that both the anonymous and non-anonymous groups provided more non-directive higher concerns (ND_HOC) feedback than other feedback types (see figure 2). Besides, the number of feedback items that was processed and not processed for ND_HOC was almost similar (see figure 4) in both conditions. This

might not be an issue of anonymity, but rather, assessors in both conditions may be able to identify HOC errors but are not able to offer concrete suggestions for change. Likewise, assessees did not process the ND_HOC feedback as these changes might require more higher level-writing skill and knowledge. This reinforces the notion that constructing higher-order feedback and acting on such feedback type is a cognitive demanding task which requires training and concrete guidelines to scaffold the core activities in the peer review process.

Regarding the effect of anonymity on writing performance, akin to the research findings by Guildford (2001) and Lu and Bol (2007), students in the anonymous condition scored higher final grades than students in the non-anonymous condition. This strengthens the belief that anonymous peer review *might* be related to improved writing skills. However, we also do not rule out the possibility of other mitigating factors e.g., the provision of peer feedback training, teacher factor etc. that could moderate the findings on writing performance. Training in providing peer feedback can increase peer feedback quality, and consequently, students' learning gains in peer review (Carless & Boud, 2018; Huisman et al., 2018). Hence, the findings on higher final scores in writing performance students in the non-anonymous condition need to be interpreted with caution. Overall, the core findings seem to indicate that anonymity could be instrumental to enhance the effectiveness of online peer review process of second-language writing courses. Taken together, the peer review process not only embodies the characteristic constructs of ZPD but also contains aspects of cognitive apprenticeship which situates the acquisition of skills and knowledge in the social and functional context of use (Collins, Brown, & Newman, 1989). Cognitive apprenticeship advocates the *between-people scaffolding* (Pea, 2004) where students can observe, enact and practice the tacit processes with help from the teacher, experts and from fellow learning partners. Next, peer review also accentuates the fading of the instructional scaffold (Pea, 2004) where peer feedback afforded learners greater agency over their own cognitive activities. This resonates with Perkins's (1993) argument that the very act of working or collaboration amounts to a kind of cognitive scaffold that would make it difficult for the individual to lose his or her place in the process. And importantly, anonymity could induce a more objective disposition and critical thinking before learners accept or reject the feedback (Gielen et al., 2010; Huisman, Saab, Van Driel, Van den Broek, 2017).

6. Limitation and Implications for Future Research

The present study offered some critical insights on anonymity in online peer review. However, the findings should be interpreted with caution owing to its inherent limitations. First of all, the participants were predominantly male which makes extrapolation of the results to other groups of students more difficult. Besides, participants were all Dutch students of Commerce and Marketing. According to Rienties, Luchoomun and Tempelaar (2014), Dutch students seem to do well in student-centred educational settings where Hofstede's cultural dimensions of low power-distance, weak uncertainty avoidance and femininity are the norm. This implies that Dutch students are used to be treated as equals, to have their opinions taken seriously, and would rather care for their peers than strive for the highest marks, all of which are important elements in the peer review process. These insights suggest that anonymity could have a different effect on the peer review process of students of other cultures and disciplines. It might be possible that the results in masculine

cultures that score high in power-distance and uncertainty avoidance show different outcomes.

The results of the study raise several other interesting research questions. First of all, it would be necessary to investigate the effect of anonymity on the long-term writing performance. Cho et al. (2006) claimed that directive comments may lead to changes in the text, but not to changes in a student's writing behaviour. This is because students receiving directive feedback might conveniently follow the provided suggestions whereas students receiving feedback in a non-directive way have to think about the situation and solve the problems themselves which kindles reflective learning. On the self-same note, a qualitative study into the effects of anonymity might lead to a more in-depth understanding of the effect of anonymity on students in the different phases of the peer review process. Rotsaert, Panadero, and Schellens (2018) explored the use of anonymity on assessors as an instructional scaffold for transition into a non-anonymous peer feedback setting. Their study found that students' attribution on the importance of anonymity decreased over time and the quality of the peer feedback was comparable in both anonymous and non-anonymous setting.

This study also raises the question whether anonymity has more effect on assessees or on assessors. The focus of the present research questions was on anonymity in the peer review process, and specifically, the impact it had on assessees. The effect of anonymity on assessors might play a role in the difference in the final grades between students who gave and received feedback anonymously and their peers in the non-anonymous group. Anonymous assessors who were less inhibited in rendering feedback of higher-order concerns (directive or non-directive) might increase in critical thinking skills, as compared to their non-anonymous counterparts over time. Likewise, the peer review process enables assessors to develop confidence - seeing how their peers perform, plus they might use their peers' work as a source for ideas and vocabulary for their own writing. As such, it would be interesting to investigate how anonymity in online peer review might affect assessors' writing performance. On this note, it would also be necessary to verify the findings of this study with a within-group design to establish the differences across the levels of the independent variable(s). Within-group design would have afforded the control of extraneous variables such as subjects' mood, gender, race, age etc.

Notwithstanding the possible limitations, we believe that the findings of this empirical study provided some useful insights into the characteristics of online peer review in a second-language writing course. It showed that anonymity had a positive effect on the three main activities defining the online peer review process. The core findings could inform the design and facilitation of the peer review process in both online or non-online learning setting.

References

Ajjawi, R., & Boud, D. (2017). Researching feedback dialogue: An interactional analysis approach. *Assessment and Evaluation in Higher Education, 42*(2), 252–265. https://doi.org/10.1080/02602938.2015.1102863

Berg, E. C. (1999). The effect of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing, 8*(3), 215-241.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315-1325. DOI: 10.1080/02602938.2018.1463354

Chi, M.T.H. (1997). Quantifying qualitative analyses of verbal data: a practical guide. *The Journal of the Learning Sciences, 6*(3), 271-315.

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

Cho, K., Schunn, C.D., & Charney, D. (2006). Commenting on writing: typology and perceived helpfulness of comments from novice peer reviewers and subject matter experts. *Written Communication, 23*(3), 260-294.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*, (pp. 453-494). New Jersey: Lawrence Erlbaum Associates.

Coté, R. A. (2014). Peer feedback in anonymous peer review in an EFL writing class in Spain. *GiST Education and Learning Research Journal, 9*, 67-87.

de Guerrero, M. C. M., & Villamil, O. S. (2000). Activating the ZPD: mutual scaffolding in L2 peer revision. *The Modern Language Journal, 84,* 51-68.

Ekşi, G. Y. (2012). Peer review versus teacher feedback in process writing: How effective. *International Journal of Applied Educational Studies, 13*(1), 33-48.

Ge, S., & Er, N. (2005). An online support system to scaffold real-world problem solving. *Interactive Learning Environments, 13*(3), 139-157. doi: 10.1080/10494820500382893

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.

Guilford, W. H. (2001). Teaching peer review and the process of scientific writing. *Advances in Physiology Education, 25*(3), 167-175.

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.

Huisman, B., Saab, N., van Driel, J., & van den Broek, P. (2018). Peer feedback on academic writing: undergraduate students' peer feedback role, peer feedback perceptions and essay performance. *Assessment & Evaluation in Higher Education, 43*(6), 955-968. DOI: 10.1080/02602938.2018.1424318

Johnson, R. (2001). The next frontier of the student-centered classroom: teaching students to recognize quality writing through the use of peer evaluation opinion paper. *Journal of Education for Business, 70*(6), 357-361.

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: a cognitive perspective. *Learning and instruction, 20*(4), 344-348.

Lee, L. (2008). Focus-on-form through collaborative scaffolding in expert-to-novice online interaction. *Language Learning & Technology, 12*(3), 53-72.

Liou, H. C., & Peng, Z. Y. (2009). Training effects on computer-mediated peer review. *System, 37*(3), 514-525.

Lu, R., & Bol, L. (2007). A comparison of anonymous versus identifiable e-peer review on college student writing performance and the extent of critical feedback. *Journal of Interactive Online Learning, 6(*2), 100-115.

Lu, J., & Law, N. (2012). Online peer assessment: effects of cognitive and affective feedback. *Instructional Science, 40*(2), 257-275. doi: 10.1007/s11251-011-9177-2

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. doi:10.1016/j.jslw.2008.06.002

Min, H. T. (2005). Training students to become successful peer reviewers. *System, 33*(2), 293-308.

Min, H, T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing, 15*(2), 118-141.

Nelson, M., & Schunn, C. (2009). The nature of feedback: how different types of peer feedback affect writing performance. *Instructional Science, 37*(4), 375-401.

Perkins, D. N. (1993). Person-plus: a distributed view of thinking and learning. In G. Salomon (Eds.), *Distributed Cognitions: Psychological and Educational Considerations* (pp. 88-110). Cambridge: Cambridge University Press.

Parr, J., & Timperley, H. (2010). Feedback to writing, assessment for teaching and learning and student progress. *Assessing Writing, 15*(2), 68-85.

Paulus, T.M. (1999). The effect of peer and teacher feedback on student writing. *Journal of second language writing, 8*(3), 265-289.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences, 13*(3), 423–451.

Raes, A., Vanderhoven, E., & Schellens, T. (2015). Increasing anonymity in peer assessment by using classroom response technology within face-to-face higher education. *Studies in Higher Education, 40*(1), 178-193. doi: 10.1080/03075079.2013.823930

Rienties, B., Luchoomun, D., & Tempelaar, D. (2014). Academic and social integration of Master students: a cross-institutional comparison between Dutch and international students. *Innovations in Education and Teaching International, 51*(2), 130-141. doi: 10.1080/14703297.2013.771973

Rollinson, P. (2005). Using peer feedback in the ESL writing class. ELT journal, 59(1), 23-30.

Rotsaert, T., Panadero, E., Schellens, T., & Raes, A. (2017). "Now you know what you're doing right and wrong!" Peer feedback quality in synchronous peer assessment in secondary education. *European Journal of Psychology of Education, 33*(2), 1-21. doi: 10.1007/s10212-017-0329-x

Rotsaert, T., Panadero, E., & Schellens, T. (2018). Anonymity as an instructional scaffold in peer assessment: its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills. *European Journal of Psychology of Education, 33*(1), 75-99.

Saeed, M. A., & Ghazali, K. (2017). Asynchronous group review of EFL writing: Interactions and text revisions. *Language Learning & Technology, 21*(2), 200-226.

Stahl, G. (2005). Group cognition in computer-assisted collaborative learning. Journal of Computer Assisted Learning, 21(2), 79–90.

Storch, N. (2019). 8 Collaborative Writing as Peer Feedback. Feedback in Second Language Writing: *Contexts and Issues,* 143.

Strijbos, J. W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction, 20* (4), 291-303.

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and instruction, 20*(4), 339-343. doi: 10.1016/j.learninstruc.2009.08.003

Tseng, S.C, & Tsai, C. C. (2007). Online peer assessment and the role of the peer feedback: a study of high school computer course. *Computers & Education, 49*(4), 1161-1174.

Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupil scan assess their peers anonymously? A quasi-experimental study. *Computers & Education, 81*, 123-132. doi: 10.1016/j.compedu.2014.10.001

Van der Pol, J., Van den Berg, B. A. M., Admiraal, W. F., & Simons, P. R. J. (2008). The nature, reception, and use of online peer feedback in higher education. *Computers & Education, 51*(4), 1804-1817.

Van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: the role of interpersonal variables and conceptions. *Learning and Instruction, 20*(4), 280-290.

Van Popta, E., Kral, M., Camp, G., & Martens, R. L. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review, 20*, 24-34.

Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, 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.

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes*. Cambridge: Harvard University Press.

Wu, W, C. V., Petit, E., & Chen, C. H. (2015). EFL writing revision with blind expert and peer review using a CMC open forum. *Computer Assisted Language Learning, 28*(1), 58-80. doi: 10.1080/09588221.2014.937442

Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005-2014). *Language Teaching, 49*(4), 461-493. doi: 10.1017/S0261444816000161

Zhu, Q., & Carless, D. (2018). Dialogue within peer feedback processes: Clarification and negotiation of meaning. *Higher Education Research & Development, 37*(4), 883–897.