Content list available at http://ijltr.urmia.ac.ir

Iranian Journal of Language Teaching Research

ORIGINAL ARTICLE

Urmia University

Effects of Scaffolded Peer Review Training on Revision Quantity and Quality in Foreign Language Writing

DO Thi Bich Thuy [a, *]

[a] *VNU University of Languages and International Studies, Vietnam National University, Hanoi, Vietnam*

ABSTRACT

While several research studies have investigated the impact of peer review training (PRT) on writing quality of L2 students, effects of PRT on revision quantity and quality remain vague. The purpose of this study was to compare revision quantity and quality before and after a scaffolded PRT following Anderson's proceduralization theory. Twenty French-major Vietnamese college students were trained for two months to become acquainted with peer review processes: detection, diagnosis and correction. Texts and feedback from pairs of students were collected and analysed in terms of cognitive revision processes. Mann-Whitney U tests showed significant effects of the scaffolded PRT on detection quantity, diagnosis quantity, diagnosis quality and correction quality. Students made more error detections, more compliments, more diagnoses, and more macro-level detections but fewer micro-level detections in the first draft of their partner's text after PRT. They generated better diagnoses in the first draft of their partner's text and better corrections in the final draft of their text after training. The author recommends that language teachers continue to offer scaffolded PRT to improve students' writing competence.

Keywords: peer review training; revision quantity; revision quality; proceduralization theory; L2 writing; French as a Foreign Language

ARTICLE HISTORY

Received: 19 June 2020 **Revised version received:** 21 Jan. 2021

Accepted: 25 May 2023 **Available online:** 1 July 2023

* Corresponding author: VNU University of Languages & International Studies, Vietnam National University, Hanoi, Vietnam
Email address: thuydtb1976@vnu.edu.vn

10.30466/ijltr.2023.121330

Introduction

Peer review (PR) is a collaborative classroom practice in which students work in groups of two, three or four and are asked to revise the text of a peer in order to improve it (Hansen & Liu, 2005). Peer assistance studies have focused on the teaching of English (to native speakers) since the early 80s, and on the teaching of English as a second or foreign language since the 90s. The results of these studies vary depending on the participants, the experimental procedure, and the data collection; but the latest research claims that PR is beneficial for the teaching and learning of writing (Do, 2020; Lu et al., 2023; Tian & Zhou, 2020; Uymaz, 2019; Wu & Schunn, 2021).

It is important to recognise the role of training or scaffolding the PR processes for the effectiveness of writing instruction (Bouwer et al., 2017; Chang, 2015; De Smedt & Van Keer, 2018; Do, 2020). Performing appropriate revisions of a peer's text is not a skill for which most students have been prepared. It is therefore improbable to expect that they will have the ability to detect problems in their peer's text, to give satisfactory explanations about the nature of these errors, or to help their peer to successfully correct these errors. If teachers want their students to skillfully revise their peer's text, they need to clearly structure the peer review training (PRT) in the writing lessons.

The present study aims to fill the knowledge gap about the impact of a scaffolded PRT on revision quantity and revision quality in French as a foreign language (FFL). Its purpose is to elucidate the effects of PRT on the quantity and quality of the detections, the diagnoses and the corrections made by FFL students in a writing class, using a range of quantitative measures. The revision model of this study is based on the model of revision processes described by cognitive psychologists Butterfield et al. (1996). The PRT in this study was developed from the social constructivism concept of "Scaffolding" and Anderson's proceduralization theory.

Literature review

Theoretical framework

Revision processes of Butterfield et al. (1996)

The present research adopted the model of revision processes provided by Butterfield et al. (1996). It includes two components. (1) The "Task environment" encompasses the rhetorical and pragmatic dimensions of the production (theme, audience and importance) and the representation of the written text during the revision (format, genre, lexical units, syntactic units, propositions, and essential ideas). (2) The "Cognitive / metacognitive" system is divided into working memory and long-term memory. Working memory is where the controlled processing of information takes place. The reviewer develops a representation of the rhetorical problems and a representation of the text produced so far. He/she ***detects*** problems by perceiving a gap between the written text and the initial required representation. Then, he/she ***diagnoses*** those errors by determining their nature. Finally, he/she helps the writer to ***correct*** errors by modifying or creating strategies for reviewing the written text. Long-term memory is divided into two levels. The cognition level provides the storage of knowledge (referential, linguistic and self-evaluative), strategies required by the revision processes (rereading a difficult passage, going back, comparison of several possibilities of revision) and the representation of the text being revised. The metacognition level stores models of knowledge and understanding of strategies. Confronting these models with the represented text allows the reviser to understand and analyse the strategies and knowledge he/she used to produce his/her text.

According to Alamargot and Chanquoy (2002, p. 54), this model of Butterfield et al. (1996) is "the most complete revision model" because it integrates revision operations into a general cognitive system composed of long-term memory and working memory. In addition, it specifies the revision operations (detection, diagnosis, correction) that are managed by the working memory and that can be worked on in a language class.

Social constructivism concept of Scaffolding

According to Wood et al. (1976), scaffolding is a kind of assistance given by an expert (an adult, a teacher, more capable peer) to a novice (a child, a learner). Children internalize information around them through social interactions with experts but to be useful, those interactions must occur in their "zone of proximal development" (ZPD). Vygotsky (1978, p.86) defined ZPD as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem-solving under adult guidance, or in collaboration with more capable peers". ZPD is a gap between what a person can accomplish alone and what this person can do with assistance (Lantolf, 2000). Scaffolding means (1) recruiting interest in the task, (2) simplifying the task, (3) maintaining pursuit goals, (4) marking critical features and discrepancies between what has been produced and the ideal solution, (5) controlling frustration during problem solving, and (6) demonstrating an idealised version of the act to be performed (Wood et al., 1976).

In a language class, the expert is usually the teacher who provides scaffolded assistance to students. Teachers should gradually reduce the quantity as well as the quality of their support to learners and encourage students to perform the task independently. Many later studies reported that scaffolded help may also come from peers (De Guerrero & Villamil, 2000; Samana, 2013; Tian & Zhou, 2020).

Anderson proceduralization theory

Anderson (1983) distinguishes two kinds of knowledge: declarative knowledge and procedural knowledge. Declarative knowledge "consists of information that can be described verbally" (Masson, 1990, p. 223) such as rules to conjugate a verb in French. Procedural knowledge represents "how to do things" and consists of condition-action (IF-THEN) pairs, called productions. The proceduralization process is "a transition from control by declarative knowledge to control by procedural knowledge" (Masson, 1990, p. 224). In second language acquisition, this process happens when students verbally learn declarative knowledge, then practice with declarative knowledge in their working memory. Much practice will eventually generate skilled performance or productions that are activated without declarative procedures. This theory will be dealt with in more detail in section 3.4.

Previous studies investigating effects of PR on revision types and reviewer stances

Effects of PR on revision types

Some studies were conducted to measure the effects of PR on revision types in second language writing proficiency (Arnold et al., 2012; Berg, 1999; Min, 2006; Paulus, 1999; Tsui & Ng, 2000; Yang et al., 2006). The majority of the studies measuring the impact of PRT on revision competence placed a great deal of importance on scaffolding the PR. In the research by Berg (1999), students examined several professional drafts with comments from reviewers and viewed one video clip as an example of an unsuccessful peer response interaction and one video clip as an example of a successful one. Appropriate language and techniques for giving feedback had

been modelled and discussed, and the importance of focusing on the ideas and organization of the essay was highlighted in the PR form (Paulus, 1999). Training activities became increasingly procedural with demonstrations of PR strategies, checklists, and personalized follow-up with reviewers (Min, 2006; Yang et al., 2006).

To analyse the impact of collaborative revision, all the above-mentioned research examined students' drafts. Berg (1999) also used verbalizations of students during their revisions. The size of the corpus differed greatly from one research to another. Berg's corpus was the largest (92 drafts) and Paulus's the smallest (22 drafts). Min (2006) investigated the impact of two drafts, before and after the PRT, while the other researchers worked on texts collected after training.

For data processing, the majority of researchers used the taxonomy of revision changes provided by Faigley and Witte (1981) which divides revisions into two major categories: surface changes (changes that do not affect the meaning) and meaning changes (changes that affect the concepts and meaning by bringing new information to the text). Surface changes include formal changes (editing) and meaning-preserving changes (paraphrasing the original concepts in the text by making them implicit or explicit, without altering the meaning). Meaning changes include microstructure changes (simple adjustments or elaborations of existing text which do not affect the overall summary, gist, or direction of the ideas in the text) and macrostructure changes (altering the text's overall direction and gist). Paulus (1999) offered the most complete application of Faigley and Witte's (1981) taxonomy with an analysis of four types of revisions (formal surface changes, meaning-preserving surface changes, meaning changes at the microstructure level and meaning changes at the macrostructure level); Min (2006) with three types of revisions (meaning changes at the microstructure level, meaning changes at the macrostructure level and surface changes); Arnold et al. (2012) with three types of revisions (formal changes, meaning-preserving changes and meaning-developing changes); Yang et al. (2006) and Chang (2015) with two types of revision (meaning changes and surface changes); and Berg (1999) with only one type (meaning changes). Yang et al. (2006) also identified the feedback as usable or not usable and classified the revisions into one of three categories: success, failure, or mixed effect (categories adapted from Conrad & Goldstein, 1999).

In terms of results, all researchers found that peer-initiated revisions affected the meaning rather than the form of the text. For example, semantic revisions accounted for two-thirds of the peer-influenced revisions in the study by Paulus (1999). Similarly, Berg (1999) found that the experimental group (benefiting from PRT) made significantly more semantic revisions than the control group (without training). Yang et al. (2006) found that participants with feedback from peers made more semantic revisions than participants with feedback from the teacher. Moreover, Min (2006) revealed that 56% of peer-influenced revisions were macrostructure changes. Arnold et al. (2012) reported that students worked more with peers when they needed to change the content of their text.

It is important to note that previous research was limited to the taxonomy of revisions provided by Faigley and Witte (1981) which allows teachers to unobtrusively record revision changes, especially in terms of correction operations. But given that this research method does not enable researchers to describe revision as a process, it will not be used as an analytical model for the present study. Revision, defined as a complex mental activity, should be analysed in terms of cognitive operations such as detecting, diagnosing and correcting errors (Butterfield et al., 1996).

Effects of PR on reviewer stances

Several studies were undertaken to classify reviewer stances during peer feedback interactions in a social approach investigating how social environment influences human behaviour (Arnold et al., 2012; Chang, 2015; Lockhart & Ng, 1995; Min, 2008).

Lockhart and Ng (1995) examined oral responses and described four types of reviewer stances: authoritative (alerting the writers about potential problematic parts), interpretive (giving opinions), probing (checking writers' intentions) and collaborative (making specific suggestions). They concluded that students who adopted probing and collaborative stances tended to give information rather than opinions and therefore provided the most useful feedback to their peers because these two stances provided students "a fuller understanding of the writing process" (Lockhart & Ng, 1995, p. 606).

Min (2008) reported on the positive impact of PRT on reviewer stances in an EFL writing class in Taiwan. She used a modelling of a four-step procedure during her PRT: clarifying writers' intentions, identifying problems, explaining the nature of problems, and making specific suggestions. According to Min (2008, p. 290), "each step, respectively, represented the most salient characteristic of readers that assumed probing, prescriptive, tutoring, and collaborative stances in the literature". Before training, the prescriptive stance (identifying problems) was prevalent but after training, there was an increase in the collaborative stance (making specific suggestions).

Chang (2015) showed that teacher modelling (where the teacher demonstrates the desirable behaviours of a reviewer) improved the feedback of student reviewers by increasing their audience awareness, and, thus, providing more helpful PR. In her study, the teacher modelled the collaborative stance to students: "Whenever a problem was identified, she explained it and offered concrete examples and revision-oriented suggestions to improve it" (p. 10). Paired t-tests showed significantly higher percentages of global feedback and collaborative stance after training, meaning that more revision-oriented suggestions from reviewers were incorporated into the second version of the writers.

Taken together, previous studies on reviewer stances revealed that the collaborative stance—when students are able to give specific suggestions which can be later used by their peer in the revised text—is the most effective in peer interactions. The collaborative stance or correction process is also the highest level of the revision operations (Compare-Diagnose-Operate) in the working memory (Bereiter & Scardamalia, 1987; Hayes et al., 1987). The four-step procedure in Min's (2008) study and the collaborative stance modelling in Chang's (2015) are quite similar to the revision processes (detection, diagnosis, correction) models of cognitive psychology which was used as the theoretical framework in this study. Clarifying writers' intentions and identifying problems correspond to the error detection process; explaining the nature of problems is equivalent to the error diagnosis process; and making specific suggestions matches the error correction process. However, Min (2008) did not employ any parametric or nonparametric statistics to detect changes after PRT. Furthermore, Min (2008) mainly investigated reviewer stance changes in terms of quantity (which stance is dominant after PRT) and not in terms of quality (which stance is the most improved after PRT).

Research questions

The current study thus proposes a cognitive approach of investigating the effects of scaffolded PRT on revision quantity and quality of FFL learners. A PRT experiment was carried out in a class of twenty freshmen in a French department. Forty pre-training composition first and final drafts and forty post-training composition first and final drafts were collected. This study was guided by the following research questions:

1. What are the effects of scaffolded PRT on error detection, diagnosis and correction quantity of FFL students?

2. What are the effects of scaffolded PRT on error detection, diagnosis and correction quality of FFL students?

Method

Participants

Participants were twenty first-year students at the University of Languages and International Studies, Vietnam National University, Hanoi. There were 19 females and 1 male, with an average age of 18. Their mother tongue was Vietnamese. All were French majors and had passed the National French Test administered by the Ministry of Education and Training in Vietnam before being admitted to the university. Their French proficiency was around B1 level for linguistic knowledge and reading, but only at the A2 level for the listening, writing and speaking skills. The writing course aimed to help first-year French students achieve A2+ (CEFR) proficiency and focused on narrative texts of 180-200 words (cf. Appendix A for an example of a writing topic). This class met once a week for 12 weeks (each class session lasting 100 minutes).

Writing cycle

The writing cycle in use in the French department was traditional: brainstorming → draft → teacher's written feedback → collective correction session. The writing cycle was longer in the PR class: brainstorming → first draft → pair's oral and written feedback (both in Vietnamese and French but mostly in Vietnamese) → final draft → teacher's written feedback on the final draft and on the pair's feedback → collective correction session. This is quite similar to Paulus' (1999) writing cycle and not as complicated as Tsui and Ng's (2000) or Min's (2006) models. Participants produced five compositions with PR working sessions. All the final drafts were graded to hold students accountable, but, given that PR practice was a new and challenging task, reviewers' feedback received only written and oral comments from the teacher but was not marked.

A collective correction session, lasting around 40 minutes, took place when drafts with the teacher's written feedback were given to students. The teacher provided her feedback on ideas, coherence, vocabulary and grammar and focused on errors due to interference between Vietnamese and French (for example the difference between *aller* and *venir*, *connaître* and *savoir*...), which even more competent students were concerned about. A collective correction sheet was elaborated for each writing task, listing the most common errors that could not be corrected by

the students themselves and errors that could be useful for all writers. Students in class discussed and suggested ways to correct these errors. They revealed later during the last lesson that they had learnt a lot from these collective correction sessions. Students were asked to sit in pair writing groups so that reviewers could read the teacher's comments about reviewers' feedback on the first draft. Moreover, teacher feedback was provided not only on errors made by writers but also on reviewers' work. For example, when a reviewer commented that "il ne fait pas froid à Lyon en Juin" – "it is not cold in Lyon in June", she was complimented on her feedback; when a reviewer wrote: "ce que tu as raconté ne m'intéresse pas" – "what you said did not interest me", she was told not to be so rude.

PR pairs

The twenty participants were divided into 10 pairs according to the pre-test results[1]. Three very good students (scoring from 8 to 9.5/10 points in the pre-test) were paired with 3 weak ones (from 1 to 3.4), 3 good students (from 7 to 7.9) with 3 average ones (from 3.5 to 5.4) and 4 quite good students with 4 quite good ones (from 5.5 to 6.9). All the pairs stayed the same during the entire semester. This method was employed because the teacher hoped that more competent students could assist weaker students in improving their writing. The choice to form PR groups of 2 students is due to two reasons. According to Arnold et al. (2012), Hu (2005) and Paulus (1999), discussions tend to be more intensive in pairs. In addition, Hu (2005, p.330) suggests that "students in dyads are more comfortable sharing each other's work". De Smedt and Van Keer (2018, p. 348) recommended that future studies should "opt for pair writing" for a more "structured application of peer assistance".

PRT and practice

The PRT and practice included three stages based on Anderson's proceduralization theory. The first stage involved conceptualization (developing the concept) of the PR processes (4 class sessions), followed by practice of PR with declarative knowledge (6 class sessions) and then proceduralization, i.e., the automation of the PR process without declarative knowledge (2 class sessions) (cf. PRT Calendar in Appendix B). The conceptualization stage was the key to success of the training because PRT was totally unfamiliar to each student. The first stage was divided into three steps: problem-situation (writing task) solving, explaining procedures used by students during their PR session, and elaborating PR processes using a student draft (composition 1). Students were assigned to write the first composition, then to work in pairs and revise their partner's draft. Then they were asked to explain orally to the class the procedures they had employed in their PR processes (the types of errors they had focused on, the quantity of rereading they did and their reading strategies...). Finally, the teacher helped the class to develop the concept of PR processes (see Table 1) using a projector. She modelled how to make detections, diagnoses and corrections on an anonymous draft (not from the PR class) and typed the feedback directly on the electronic draft. Feedback was given on the textual aspects and then the linguistic aspects. Like Hu (2005), modelling was demonstrated on a full text, not on an extract, to better analyse the task completion, the structure and the coherence. The teacher wanted the students to experience the PR before her modelling, so with her help, they could conceptualize the collaborative revision processes from their practices and better understand how to do it in an efficient way.

Table 1 shows PR processes developed during the writing class and used in the PRT.

Table 1
PR Processes (Adapted from Butterfield et al., 1996)

Revision process	Revision tasks
1. Error detection	The reviewer checks the length of the text, looks at the overall structure of the text, checks the suitability of the text for the task completion. The reviewer assesses the author's ability to communicate (ability to tell and describe facts and feelings, ability to affect the reader) and the coherence of the text. The reviewer provides a summary of his feedback on textuality. The reviewer checks the quality of the language (syntax, vocabulary, and spelling) of the text. The reviewer summarizes his feedback on the language.
2. Error diagnostic	Both peers discuss to diagnose detected errors. Sources of help for the diagnosis can be dictionaries, online resources, teacher or friends.
3. Error correction	Both peers discuss to correct detected errors. Sources of help for the correction can be dictionaries, online resources, teacher or friends. The writer completes corrections to the text.

In the second stage, students practiced PR processes with declarative knowledge (with PR checklists, table of PR processes, and sheets of advice on how to give and receive feedback) during 6 class sessions. The teacher demonstrated to them PR processes on two more students' drafts (compositions 2 and 3). She modelled how to detect errors or give compliments, how to explain the nature of each error to their partner in oral and written forms and how to correct errors using a French-French dictionary and online resources. A PR working session lasted around 40 minutes. Pairs read the draft, detected and diagnosed errors, discussed with the writers how to make the corrections, and then wrote their feedback on the draft. Students provided their pairs with feedback 3 times (compositions 2-4) in the second stage of the proceduralization.

In the last stage of the proceduralization, students practiced PR on the fifth composition without declarative knowledge (no demonstration session, no PR checklist, and no sheets of advice on how to give and receive feedback). The teacher wanted to verify if students achieved the automation of the processes.

PR tools

PR tools consisted of sheets of advice on how to give and receive feedback and customized PR checklists. They were employed at stage 2 of the PRT when students practiced PR processes with declarative knowledge.

Sheets of advice on how to give and receive feedback

Two sheets of advice on how to give feedback and on how to receive feedback were provided. They were adapted from Lundstrom's sheets (Lundstrom & Baker, 2009). They aimed to give students strategies for successful interactions with pairs. For example: 'Focus on the structure and the meaning of the text, not only on grammar, spelling or punctuation"; "Be specific – don't just say "it is good". Tell the writer what is good"; "Ask questions if there is something you do not understand". In these sheets, examples of good and bad feedback as well as examples of good and bad reactions to feedback were also included.

The teacher asked students to give compliments before moving on to criticism. The purpose was to give constructive feedback in a way that did not discourage their pairs and to specify what was wrong in the text.

Customized PR checklists

A PR checklist was created by the students and the teacher for each composition (cf. Appendix A). The checklist was divided into 2 parts: questions on textuality (between six and nine questions) and questions on language (three questions). The first dealt with five aspects: the writing situation (i.e., who writes, to whom, when, where?), the type of text (i.e., the forms of a familiar letter and an unfamiliar letter), the relevance and richness of ideas, the structure and the coherence. The second part concerned syntax, vocabulary and spelling. While the questions on language were the same in all checklists, the questions on textuality varied according to writing instructions.

It should be noted that textuality questions accounted for two-thirds, or three-quarters of the PR questions. Students were encouraged to first review ideas before moving to local correction. It was a deliberate choice of the teacher to put the ideas of a text in the foreground.

Data analysis

The data included forty first and final drafts of the pre-training composition and forty first and final drafts of the post-training composition, as well as written pair feedback. Text analysis was conducted to investigate the effects of trained PR on revision quantity and quality.

Detection, diagnosis and correction coding

One independent rater and the teacher carefully read student reviewers' comments in response to the 1st drafts of the 1st composition (prior to PRT) and those to the 1st drafts of the 5th composition (after PRT). Then, they checked how student writers integrated these comments in the 2nd version of their 1st and 5th compositions. They tried to identify the corresponding revision process and then coded each feedback provided by the reviewers and each feedback used by the writers as detection, diagnosis or correction. Detection is the process where reviewers found a positive point or a mistake in their pair's text. Detection can be a compliment or a criticism; it can pertain to the macro-text level or the micro-text level (e.g. *Cher parents* is criticism on the micro-text level). Diagnosis is the process where reviewers check writers' intentions or explain the nature of errors made by writers in Vietnamese or French (e.g. *adjective agreement).* Correction refers to the process where writers correct errors detected by reviewers and integrate these corrections into the final draft (e.g. *Chers parents*).

Detection, diagnosis and correction quantity

The total number of detections, criticisms, compliments, macro-level detections, micro-level detections, diagnoses and corrections were calculated for pre-training and post-training compositions. The total number of errors made by a writer in each composition was also counted. Then the total number of detections, criticisms, compliments, macro-level detections, micro-level detections, diagnoses and corrections were divided by the total number of errors made by a writer in each composition before and after training to determine the percentage of detection quantity, criticism quantity, compliments quantity, macro-level detection quantity, micro-level detection quantity, diagnosis quantity and correction quantity before and after PRT.

Detection, diagnosis and correction quality

To measure detection, diagnosis and correction quality, each detection, diagnosis and correction was evaluated by two raters as successful or unsuccessful. A successful detection meant the problem pointed out by a pair was correctly identified. An unsuccessful detection meant the error pointed out by a pair was not an error. A successful diagnosis assessed good explanations of the nature of problems; an unsuccessful diagnosis assessed bad explanations of the nature of problems. A successful correction indicated an error that was correctly changed by a writer; an unsuccessful correction indicated an error that was incorrectly changed by a writer.

To determine detection quality, the number of successful detections was divided by the number of detections made by a reviewer in the first draft. To assess diagnosis quality, the number of successful diagnoses was divided by the number of diagnoses made by a reviewer in the first draft. To calculate correction quality, the number of successful corrections was divided by the number of corrections made by a writer in the final draft.

Inter-rater reliability

One independent rater and the teacher coded detections, diagnoses and corrections. After a "norming" session where they discussed and decided together the coding of peer feedback on 3 first drafts, they proceeded to rate the remaining comments individually. If there was a difference in coding, they would exchange opinions to come to an agreement. Inter-rater reliability was calculated resulting in a strong agreement between raters (Cronbach's alpha in SPSS version 20 = .85 for detection, = .86 for diagnosis and = .83 for correction).

Non-parametric statistics

A Shapiro-wilk test in SPSS version 20 showed that the data did not meet the assumptions for parametric statistics. A Mann-Whitney U test was conducted for detection, diagnosis and correction quantity and quality. The p-value chosen for this study was 0.05.

Results

Reviewers' detection, diagnosis and correction quantity before and after training

Table 2 provides information about detection, diagnosis and correction quantity before and after training.

Table 2
Mann-Whitney U Test Results for Detection, Diagnosis and Correction Quantity Before and After Training

Measures	Before training		After training		U	P
	N	Median	N	Median		
Detection/ total errors	20	.52	20	.68	95.500	.004
Criticisms/ total errors	20	.44	20	.42	180.500	.602
Compliments/ total errors	20	.03	20	.20	11.000	.000
Macro level detections/ total errors	20	.03	20	.34	12.000	.000
Micro level detections/ total errors	20	.39	20	.33	127.000	.049
Diagnosis/ total errors	20	.06	20	.13	108.000	.012
Corrections/ total errors	20	.40	20	.41	183.000	.66

As shown in Table 2, after training, reviewers made more detections: Mdn (before training) = .52; Mdn (after training) = .68; U = 95.500, p = .004 ($p < 0.05$). The difference is statistically significant. Results revealed that students made more compliments after training: Mdn (before training) = .03; Mdn (after training) = .20; U = 11.000, p = .000 ($p < 0.05$). Therefore, the difference is statistically significant. For criticism quantity, the difference between before and after training was not found to be statistically significant at the 0.05 level. Reviewers made more detections on the macro level and fewer detections on the micro level: Mdn macro level detections (before training) = .03, Mdn macro level detections (after training) = .34, U = 12.000, p = .000 ($p < 0.05$); Mdn micro level detections (before training) = .39, Mdn micro level detections (after training) = .33, U = 127.000, p = .049 ($p < 0.05$).

Similarly, reviewers made more diagnoses after training: Mdn (before training) = .06, Mdn (after training) = .13, U = 108.000, p = .012 ($p < 0.05$). Therefore, the difference is statistically significant.

However, writers did not generate more corrections after training: Mdn (before training) = .40, Mdn (after training) = .41, U = 183.000, p = .69 ($p > 0.05$). Therefore, the difference is not statistically significant.

In summary, students produced significantly more detections, more compliments, and more macro-level detections, but less micro-level detections. They made significantly more diagnoses, but no difference was found in terms of correction quantity.

Reviewers' detection, diagnosis and correction quality

Table 3 shows the results of detection, diagnosis and correction quality before and after training.

Table 3
Mann-Whitney U Test Results for Detection, Diagnosis and Correction Quality Before and After Training

Measures	Before training		After training		U	P
	N	Median	N	Median		
Successful detections/ detections	20	.82	20	.90	134.500	.076
Successful diagnoses/ diagnoses	20	.5	20	1.0	81.500	.001
Successful corrections/ corrections	20	.46	20	.81	78.500	.001

Results show that reviewers did not make better detections after training: Mdn (before training) = .82; Mdn (after training) = .90; U = 134.500, p = .076 (p >0.05). The difference is statistically not significant. They were able to give better error diagnoses after training: Mdn (before training) = .5; Mdn (after training) = 1.0; U = 81.500, p = .001 (p <0.05). The difference is statistically significant. Writers made better corrections after training: Mdn (before training) = .46; Mdn (after training) = .81; U = 78.500, p = .001 (p <0.05). Therefore, the difference is statistically significant.

Taken together, significant differences in terms of diagnosis quality and correction quality between the pre-training and post-training were found but not in terms of detection quality.

Discussion

While the impact of PR on revision types or on reviewer stances has been discussed in the literature, little is known about the effects of trained PR on revision quantity and quality. This study therefore aimed to deal with this gap.

Our findings of enhancement of revision quantity and quality are consistent with Chang (2015), Min (2008), and Yang et al. (2006). Min (2008) found an increase of the tutoring stance (explaining problems) from 11% to 23%, an enhancement of the probing stance (clarifying writer's intentions) from 11% to 21% as well as an improvement of the collaborative stance (making specific suggestions) from 17% to 29% after PRT. Reviewers in Chang's (2015) study produced more revision-oriented suggestions (collaborative stance) after training. Yang et al. (2006) found that peer feedback was more successfully integrated in the final draft than teacher's comments because pairs could discuss and give explanations about the nature of the errors in their partner's texts. However, this is not in line with Huang's (2004) and Mangelsdorf and Schlumberger's (1992) works which showed that reviewers mostly were able to detect errors but were not capable of offering specific corrections. Our encouraging result may be related to the scaffolded PRT where the teacher deliberately modelled in demonstration sessions and in collective correction sessions how to explain the nature of each error to their partner in oral and written forms and how to correct errors using French-French dictionaries and online resources. In addition, students worked in pairs for one semester; therefore, they knew their partners well and dared to ask clarification questions to diagnose errors and to give specific suggestions. Finally, improved writing competence acquired after the writing course would have allowed students to give better corrections to their partners.

Finding an increased percentage of compliments confirms the results of Chang (2015), Guardado and Shi (2007) and Min (2008) who showed that foreign language students in PR situations knew very well how to give compliments to their peers or used more socio-affective items after PRT. Some factors might favour the increase in compliments of this study. On the PR sheet (cf.

Appendix C), a space for compliments was designed. In addition, the teacher asked students to give positive feedback to their partner before moving to negative feedback. Finally, it would appear that writers made progress with writing practice.

The higher percentage of detections on the macro level and lower percentage of detections on the micro level corroborate several studies that have investigated revision (Arnold et al., 2012; Berg, 1999; Chang, 2015; Evans, 2017; Min, 2006, 2008; Paulus, 1999; Yang et al., 2006). However, it is not in line with Ranalli et al. (2018) who reported that their students made mostly local-level changes and avoided global changes when they revised their own text because of limited motivation or the time-consuming planning stage. There are two plausible explanations for our finding: (a) the majority of questions in the checklists were at the macro level and (b) the working memory might have been overloaded (Baaijen & Galbraith, 2018; Mehrpour et al., 2023; Saadat & Zahed Alavi, 2020) because of the PR, and students might choose to focus on improving macro level at the expense of micro-level detections.

In short, this research indicates that scaffolded PRT had a positive impact on students' detection quantity, diagnosis quantity, diagnosis quality and correction quality. However, detection quality and correction quantity were not enhanced. The number of successful detections increased after training, but the difference was not significant. These findings are in line with Yang et al. (2006), who found that peers provided less usable feedback than the teacher. One thing worth pointing out is that sometimes weak students used sophisticated words or sentences they had found in a book or a dictionary, and their very good partners underlined those new linguistic materials as an error without checking it. It appears that the negative image reviewers had of their partner's competence might have had negative effects on the quality of their detection (Strijbos et al., 2010). Consequently, weak writers refused to integrate the bad detections. Another reason for the quantity of unchanged corrections may be the increase of macro-level detections. After training, reviewers gave more detections on global issues, which are harder to correct than local errors.

Conclusion

In summary, the present paper shows that PRT with a great deal of scaffolding has a positive impact on students' revision quantity and quality. Using the cognitive revision processes model of Butterfield et al. (1996), this study analysed three major revision operations, namely, detecting positive and negative points, diagnosing errors and correcting errors. The scaffolded PRT in an FFL class during one semester demonstrates that students made more error detections and diagnoses, gave more positive feedback, and identified more macro-level detections but identified fewer micro-level detections. They produced better diagnoses and better corrections after training.

This study has some limitations. First of all, the corpus size is relatively small. Secondly, only written feedback was collected. Oral feedback would have given better information about the quantity and quality of diagnoses and corrections (Colognesi et al., 2020). Robert and Brunette (2016) suggested that further research could investigate the ability to "justify aloud" which may have a positive effect on the error detection competence. They found a correlation between the quantity of diagnoses and/ or corrections verbalised by revisers and the number of errors they detected (Robert & Brunette, 2016).

We recommend that additional work should be undertaken with pairs of similar writing proficiencies (Puranik et al., 2017; Strijbos et al., 2010) because the negative evaluation of the peer's competence apparently led to bad detections. We hope that our research will serve as a first

step for further studies on a wider level determining the impact of PRT on the quantity and the quality of detection, diagnosis and correction processes. Based on the findings of the current research, we suggest that teachers continue to offer scaffolded PRT to help foreign language students to improve their writing skills.

References

Alamargot, D., & Chanquoy, L. (2002). Les modèles de rédaction de textes. In M. Fayol (Ed.), *Production du langage: Traité des sciences cognitives* (pp. 45-65). Lavoisier.

Anderson, J. R. (1983). *Cognitive science series. The architecture of cognition.* Lawrence EErlbaum Associates, Inc.

Arnold, N., Ducate, L., & Kost, C. (2012). Collaboration or cooperation? Analyzing group dynamics and revision processes in Wikis. *CALICO Journal 29*(3), 431-448.

Baaijen, V., & Galbraith, D. (2018). Discovering through writing: Relationships with writing processes and text quality. *Cognition and Instruction 36*(3), 199-223. 10.1080/07370008.2018.1456431.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Erlbaum.

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

Bouwer, R., Koster, M., & Van den Bergh, H. (2017). Effects of a strategy-focused instructional program on the writing quality of upper elementary students in the Netherlands. *Journal of Educational Psychology, 110*(1), 58-71.

Butterfield, E., Hacker, D., & Albertson, L. (1996). Environmental, cognitive and metacognitive influences on text revision: Assessing the evidence. *Educational Psychology Review, 8*(3), 239-260.

Chang, C. Y. (2015). Teacher modelling on EFL reviewers' audience-aware feedback and affectivity in L2 peer review. *Assessing Writing, 25*, 2-21. 10.1016/j.asw.2015.04.001.

Colognesi, S., Vassart, C., Blondeau, B., & Coertjens, L. (2020). Formative peer assessment to enhance primary school pupils' oral skills: Comparison of written feedback without discussion or oral feedback during a discussion. *Studies in Educational Evaluation, 67*.

Conrad, S. M., & Goldstein, L. M. (1999). ESL student revision after teacher-written comments: Text, contexts, and individuals. *Journal of Second Language Writing 8*(2), 147-179.

De Guerrero, M.C., & Villamil, O.S. (2000). Activating the ZPD: Mutual scaffolding in L2 peer revision. *Modern Language Journal, 84*(1), 51-68.

De Smedt, F., & Van Keer, H. (2018). Fostering writing in upper primary grades: a study into the distinct and combined impact of explicit instruction and peer assistance. *Reading and Writing, 31,* 325-354. 10.1007/s11145-017-9787-4.

Do, Thi Bich Thuy. (2020). Effects of scaffolded peer review training on the quality of texts produced by students of French as a Foreign Language. *Electronic Journal of Foreign Language Teaching, 17*(1), 318-330.

Evans, M. (2017). *Writing in English as a foreign language within higher education in Vietnam: an investigation of the genres, writing processes and perceptions of ten Vietnamese students* (Doctoral dissertation). University of Warwick, Centre for Education Studies.

Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication, 32*, 400-414.

Guardado, M., & Shi, L. (2007). ESL students' experiences of online peer feedback. *Computers and Composition, 24*, 443-461.

Hansen, J., & Liu, J. (2005). Guiding principles for effective peer response. *ELT Journal, 59*, 31-38.

Hayes, J.R., Flower, L.S., Schriver, K.A., Stratman, J., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics: Reading, writing and language processing* (pp. 176-240). Cambridge University Press.

Hu, G. (2005). Using peer review with Chinese ESL student writers. *Language Teaching Research, 9*(3), 321-342.

Huang, M. C. (2004). *The use of process writing and Internet technology in a Taiwanese college English writing class: A focus on peer reviews* (Doctoral dissertation). The Pennsylvania State University.

Lantolf, J. P. (2000). *Sociocultural theory and second language learning*. Oxford University Press.

Lockhart, C., & Ng, P. (1995). Analyzing talk in ESL peer response groups: Stances, functions, and content. *Language Learning, 45*(4), 605-655.

Lu, Q., Yao, Y., & Zhu, X. (2023). The relationship between peer feedback features and revision sources mediated by feedback acceptance: The effect on undergraduate students' writing performance. *Assessing Writing, 56*.

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.

Mangelsdorf, K. & Schlumberger, A. (1992). ESL student response stances in a peer-review task. *Journal of Second Language Writing, 1*(3), 235-254.

Masson, M. E. (1990). Cognitive theories of skill acquisition. *Human Movement Science, 9*, 221-239. 10.1016/0167-9457(90)90004-W.

Mehrpour, S., Hoomanfard, M. H., & Vazin, E. (2023). Peer feedback accuracy in synchronous and asynchronous computer-mediated conditions in an EFL context. *Iranian Journal of Language Teaching Research, 11*(1), 97-116. 10.30466/ijltr.2023.121274.

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

Min, H. T. (2008). Reviewer stances and writer perceptions in EFL peer review training. *English for Specific Purposes, 27,* 285-305.

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

Puranik, C.S., Patchan, M.M., Lemons, C.J., & Al Otaiba, S. (2017). Using peer assisted strategies to teach early writing: results of a pilot study to examine feasibility and promise. *Reading and Writing, 30*, 25-50. 10.1007/s11145-016-9661-9.

Robert, I. & Brunette, L. (2016). Should revision trainees think aloud while revising somebody else's translation? Insight from an empirical study with professionals. *Meta, 61*(2), 320-345. https://www.erudit.org/fr/revues/meta/2016-v61-n2-meta02699/1037762ar/.

Ranalli, J., Feng, H.-H., & Chukharev-Hudilainen, E. (2018). Exploring the potential of process-tracing technologies to support assessment for learning of L2 writing. *Assessing Writing, 36*, 77-89. 10.1016/j.asw.2018.03.007.

Saadat, M. & Zahed Alavi, S. (2020). Variability in peer-peer scaffolding during writing tasks in L2 English. *Iranian Journal of Language Teaching Research 8*(3), 99-120. 10.30466 / ijltr.2020.120936.

Samana, W. (2013). Teacher's and students' scaffolding in an EFL classroom. *Academic Journal of Interdisciplinary Studies, 2*(8), 338-343.

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*, 291-303. 10.1016/j.learninstruc.2009.08.008.

Tian, L., & Zhou, Y. (2020). Learner engagement with automated feedback, peer feedback and teacher feedback in an online EFL writing context. *System, 91*.

Tsui, A. B. M., & Ng. M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing, 9*(2), 147-170.

Uymaz, E. (2019). The effects of peer feedback on the essay writing performances of EFL students. *International journal of Curriculum and Instruction, 11*(2), 20-37.

Vygotsky, L. (1978). Interaction between Learning and Development (M.A. Thesis). In *Mind in Society*. (Trans. M. Cole) (pp. 79-91). Harvard University Press.

Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*, 89-100.

Wu, Y., & Schunn, C. D. (2021). The effects of providing and receiving peer feedback on writing performance and learning of secondary school students. *American Educational Research Journal, 58*(3), 492-526. 10.3102/0002831220945266.

Yang, M., Badger, R., & Zhen, Y. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing, 15*, 179-200.

Appendices

Appendix A: PR checklist

Writing topic:

You have started your studies in a city far from your parents. Write to them to tell about your new life (studies, friends, travels, climates, difficulties ...).

To help you to review your pair's text, answer the following questions:

1. Did your pair write to his parents?

2. Did he/ she follow the form of a familiar letter (city and date, salutation, closing, signature)?

3. Did he/ she tell them about his/ her new student life (studies, friends, travels, climates, difficulties ...)?

4. Are his/ her ideas rich and interesting? Is there inappropriate information in his/ her text?

5. Did he/ she make paragraphs in his/ her text?

6. Is his/ her text clear and coherent?

7. Did he/ she make grammar mistakes?

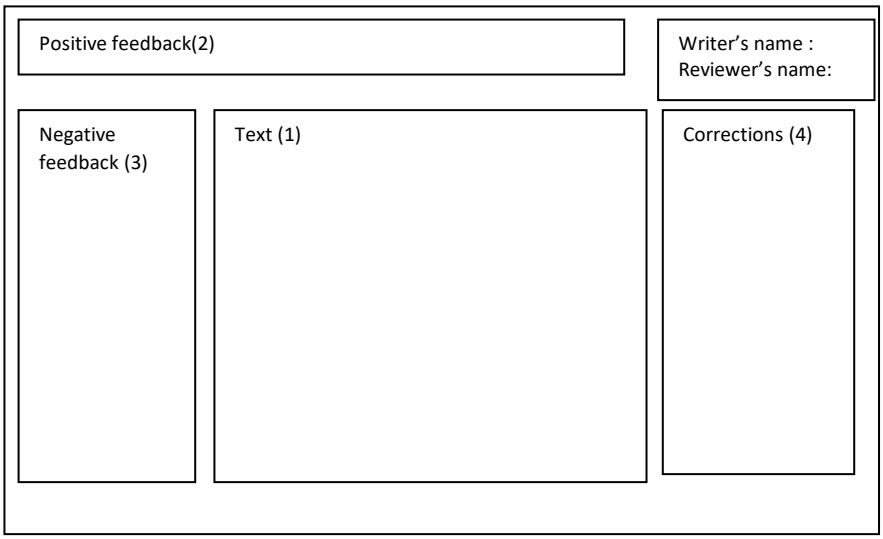
8. Did he/ she use inappropriate words? Is his/ her vocabulary rich and varied?

9. Did he/ she make spelling mistakes?

Appendix B: PRT calendar

Stages	Weeks	Activities	PR Tools
Stage 1: Conceptualization	Weeks 1, 2, 3, 4	Students wrote the pre-training text (Composition 1) and revised the first draft of their pair WITHOUT PR tools. Students wrote the final draft of the composition 1. The teacher modelled PR processes on an anonymous draft using the PR checklist for text 1. Students created PR processes table. The teacher explained the sheets of advice on how to give and receive feedback. Collective correction session of the two drafts of the composition 1.	PR sheet PR checklist for text 1 PR processes table Sheets of advice on how to give and receive feedback Collective correction sheet of the composition 1
Stage 2: Practice with declarative knowledge (with PR tools)	Weeks 5, 6	Writers wrote the composition 2. The teacher modelled PR processes on an anonymous draft (from another class) of the composition 2 using PR tools. Reviewers practiced PR on the first version of the composition 2 of their pair. Writers wrote the second version of the composition 2. Collective correction session of the two drafts of the composition 2.	PR sheet PR processes table Sheets of advice on how to give and receive feedback PR checklists for the compositions 2, 3, 4 Collective correction sheets of the compositions 2, 3, 4
	Weeks 7, 8	The same procedure was repeated for the composition 3.	
	Weeks 9, 10	The same procedure was repeated for the composition 4. But the teacher modelling was not provided.	
Stage 3: Proceduralization	Weeks 11, 12	Writers wrote the composition 5. Reviewers practiced PR on the first version of the composition 5 of their pair WITHOUT PR tools. Writers wrote the second version of composition 5. Collective correction session of the two drafts of the composition 5.	PR sheet

Appendix C: PR sheet



DO Thi Bich Thuy is a lecturer of French at the University of Languages and International Studies, Vietnam National University, Hanoi. She received a PhD. Degree from Aix-Marseille University in 2011. Her work focuses specifically on second and foreign language writing and assessment.

[1] The pre-test was an A2 writing topic. The assessment tool used was the CEFR evaluation grid for A2 level.