Levels of Equality and Gender Dispersal in Dyadic Collaborations: Does Asymmetry Work?

Shima Ghahari* , Mohadeseh Amiri, Mohamad Abbasnejad

Department of Foreign Languages, Shahid Bahonar University of Kerman, Kerman, Iran

Abstract: Although dyadic collaborations (DC) are sensitive to individual differences, the role of levels of equality (LoE) and gender disparity as potential mediators are largely unexplored. To address the research gap, this case study investigated whether symmetrical (equal and gender-matched) and asymmetrical (unequal and gender-mixed) peer interactions differ in terms of efficacy for morpho-syntactic development and learners' retrospective reflections. Forty-three young learners were randomly assigned into symmetrical (equal) and asymmetrical (unequal) pairs and into dyads of various gender dispersal (matched and mixed). The learners' behaviors and perceptions during and following the paired tasks were recorded via observation, field notes, focused-group discussion, and member-checking meetings. The results of the Mann-Whitney U-test and Friedman Test were as follows: DC was more effective than non-collaborative learning for both short-term and long-term structure and vocabulary learning. It was helpful for expert-expert and gender-matched dyads in both domains alike and for experts and novices in unequal pairs in the structure area only. It was ineffective, however, for novice-novice and cross-gender dyads. Among the recurrent themes in observation and interviews were overall preference to pair up with female partners, consensus on the advantage of DC for grammar judgment (structure) over fill-in-the-blank (vocabulary) tasks, the realization of DC as effective for promoting rapport and responsibility, experts' willingness to pair with novices or work alone, novices' tendency to work with experts, and role of personality traits as a substantial mediator. Based on the findings, several implications for research and practice are offered.

Keywords: Collaborative Learning (CL), Dyadic Collaboration (DC), Peer Feedback, Gender, Levels of Equality (LoE), Novice Learners, Expert Learners.

---

* Corresponding Author.

Authors' Email Address:

¹ Shima Ghahari (ghahary@uk.ac.ir), **² Mohadeseh Amiri** (mohadeseh25.amiri1989@gmail.com) **³Mohamad Abbasnejad** (negar25.1989@gmail.com)

Introduction

The last few decades have witnessed considerable interest in collaborative learning (CL) as a strong determinant of educational achievement. CL is typically characterized by providing a rich set of alternatives to build interaction among learners, addressing content area learning and language development needs within a normalized educational framework (Garrett & Shortall, 2002). More specifically, it increases the chances for individualized instruction, encourages consciousness-raising and attention to specific linguistic points, and enables learners to explain their points of view, experience less anxiety, gain self-confidence and learning motivation, and finally become more supportive of each other (Elola & Oskoz, 2010; Lo & Hyland, 2007; Nassaji & Tian, 2010). Much research has reported that learners who bond with classmates tend to learn more and retain the information longer, are more successful in the completion of program requirements, and enjoy higher satisfaction with the class (e.g., Dobao & Blum, 2013; Shehadeh, 2011).

Researchers have considered various settings and variables in the investigation of CL, including (a) its implications for different language macro- and micro-skills (e.g., Diab, 2010; Storch, 2011; Wigglesworth & Storch, 2009), (b) factors influencing its effectiveness (Dawson et al., 2018; Mercader et al., 2020), (c) the extent of individual engagement and involvement (Storch, 2007, 2008; Yu et al., 2018), and (d) learners' beliefs and behaviors (Chang, 2007; Huisman et al., 2019; Simonsmeier et al., 2020; Strijbos et al., 2010; Yu, 2019).

The effectiveness of dyadic collaboration (DC), as a subdivision of CL, is largely sensitive to individual differences; yet few studies to date have treated levels of equality (LoE) and gender disparity as potential mediators. The current research addressed this gap in the literature by conducting quantitative and qualitative analyses on the impact of LoE and gender diversity on language development, on the one hand, and on the learners' behaviors and perceptions during and after paired task completion, on the other. To be more specific, it investigated whether symmetrical (equal and gender-matched) and asymmetrical (unequal and gender-mixed) peer interactions differ in terms of efficacy for morpho-syntactic development [1] and learners' retrospective reflections. For this purpose, the following questions were pursued:

---

[1] Noteworthy to mention is that in this study the terms grammar, structure, and syntactic knowledge were used interchangeably, and so was the concept of vocabulary thematically equivalent to morphological knowledge (or development).

1) Does dyadic collaboration affect young language learners' morpho-syntactic development (immediate and delayed post-tests)?

2) Does levels of equality (novice-novice, expert-expert, and novice-expert) mediate the effectiveness of DC?

3) Does gender grouping (male-male, female-female, and male-female) mediate the effectiveness of DC?

4) What are the learners' actual behaviors, reflections, and suggestions during and after peer activities?

Literature Review

DC across Levels of Equality

Before addressing the issue of how levels of equality (LoE) may interact with DC, some clarification of the notion of *equality* is in order. L2 equality is defined as "the degree of control or authority over the task" (Storch, 2002, p. 127) and is classified into novice (N) and expert (E) levels. Equality differences have been debated as one of the determining factors in considering the nature and effectiveness of dyadic interactions (Yu & Lee, 2016). Van Lier (2014) asserts that more knowledgeable (expert) learners are likely to benefit from working with less knowledgeable (novice) peers in expert-novice pairs. Otha (1995) has shown that both experts and novices could benefit from dyadic interactions. In a similar vein, Watanabe's (2008) and Watanabe and Swain's (2007) studies have indicated that when engaged in collaborative patterns of interaction, learners tend to achieve higher post-test scores regardless of their partner's level. It seems, therefore, that in their studies equality differences did not necessarily affect the nature of peer assistance and L2 learning.

Controversies, however, exist in the literature as to the efficacy of DC for pairs with the highest inequality compared with pairs with the lowest ability difference (Kim & McDonough, 2008). Storch (2001) divulged that pairs with a high inequality (low and upper intermediate) were more collaborative and engaging than the other two pairs; the pairs with some degree of homogeneity (low and intermediate), however, were found to be non-collaborative, suggesting less transfer of knowledge and more missed opportunities. In contrast, Niu et al. (2018) reported that low-low pairs produced more language-related episodes and applied more diverse scaffolding strategies, while high–high and high-low pairs were more successful in task completion. Kowal and Swain (1994) documented that in a highly heterogeneous grouping (e.g., upper-middle and low), more knowledgeable learners did carry out most of the work probably because the weaker ones were too intimidated to

offer anything, were more willing to let the stronger learner do the task, or were not allowed to do the task, regardless of whether their comments were valid or not. Successful scaffolding, according to Gielen et al. (2010), requires the group members to respect and trust each other's perspectives, and this may be difficult to achieve when equality differences are too large.

Given the limited and inconsistent literature on how LoE actually interacts with the process of DC, a major undertaking in this study was to explore which form of dyadic interaction, namely equal (i.e. symmetrical; expert-expert and novice-novice) or unequal (i.e. asymmetrical; expert-novice), can be more effective in language classrooms in a typical Asian community and whether learners with divergent levels of equality derive the same educational and socio-cognitive benefits from engagement in paired activities.

DC and Gender Groupings

Research on interaction and gender differences did not start until the mid-1970s. The focus of the research has been mostly on the role of gender as an independent source of discourse patterns and conversational styles as dependent factors. In this regard, Aries's (1976) research was one of the early studies of its type investigating interactional styles of three groups, all-male, all-female, and mixed groups. Findings illustrated that males carried out more interactions than females; exercise of power, defined as the amount of talking to the group as a whole rather than to individuals, happened more often in the male-specific group; thirdly, males in the mixed group shared more about themselves and built more intimacy than in all-males; and finally, females in the all-female group were more relaxed to disclose their feelings than females in the mixed group.

In a similar vein, Tannen (1990) introduced two speech styles, namely *report-talk* and *rapport-talk*. She asserted that men tend to report-talk (public speech), which means that they are more comfortable directing their speech at large groups and talking to demonstrate their decencies or importance, while females tend to rapport-talk (private talk), which refers to the open display of similarities and shared experiences. This speech style divergence implies the importance of the concept of collaboration in females' interactions. Holmes (1992) also studied gender dispersal as a potential determinant of interactional style and concluded by formulating a set of different patterns of language use between males and females, among which were females' more uses of the affective functions of interaction, tendency to talk to individuals, and more flexibility in interactional exchanges. Finally, Pishghadam and Kermanshahi (2011) maintained that females were more willing to generate explicit and

direct feedback, while males tended to produce indirect ones; females also placed more trust in their partners' comments than males did.

The experiments reviewed above have mainly addressed such interactional facets as manners of interaction, types of feedback, as well as discourse styles of males and females, while the mediatory role of gender grouping and gender-related interactional conflicts on the DC process and outcome is fairly left untouched. Informed by these interactional variations between males and females, this study undertook to assign learners into symmetrical (gender-matched) and asymmetrical (gender-mixed) dyads to examine which composite leads to better linguistic outcomes and matches the learners' socio-cognitive needs and expectations best.

DC and Socio-Cognitive Mechanisms

Although the line of research on CL has offered insights into the socio-cognitive (e.g., Storch, 2005, 2007) and linguistic (e.g., Gass & Mackey, 2007; Wigglesworth & Storch, 2009) advancement of the individual learners, the literature is still suffering from scarce consideration of learners' perceptions and reactions towards dyadic activities. Over the last two decades, some studies (e.g., Storch, 2005; Watanabe, 2008) have inspected the effects of learners' psychological and affective variations on CL process and product. Based on their findings, learners usually disclose positive feelings and attitudes (Dobao & Blum, 2013) as well as high levels of satisfaction, motivation, and self-confidence (Shehadeh, 2011; So & Brush, 2008) as a result of DC. Garrett and Shortall (2002) examined Brazilian learners' perceptions toward paired activities and revealed that learners viewed these activities as fun, more relaxing, less anxiety-provoking, and as better learning devices than teacher-fronted assignments. Yule and Macdonald (1990) too extended the research on paired collaboration across different levels and found that all learners believed interactive turn-taking and presenting ideas, regardless of their partners' level, in a non-dominant atmosphere could assist them in solving the conflicts.

By contrast, a couple of studies in the literature have indicated that learners favor individual activities over collaborative tasks in quest for more personal ownership (Caspi & Blau, 2011), more practice time and rehearsal activities (Ghahari & Farokhnia, 2017, 2018; Strijbos et al., 2010), further opportunities to follow personal styles and schedules (Elola & Oskoz, 2010), and teachers' explicit instruction and feedback (Ghahari & Sedaghat, 2018; Ghahari & Piruznejad, 2017; McDonough, 2004; Van Gennip et al., 2010). Motivated by this inconsistency, another challenge addressed in this research is learners' perceptions and

reflections towards DC practices when paired up in heterogeneous (unequal) and homogeneous (equal) as well as gender-matched and gender-mixed dyads.

Methodology

Research Design

The study features quasi-experimental research with a pretest-treatment-post-test (immediate and delayed) design with one control and four experimental groups. Three sets of variables were involved: (a) dyadic tasks or peer collaboration as the independent variable, (b) vocabulary and structure development as the dependent variables, and (c) levels of equality and gender difference as mediators. The treatment, the independent variable, was composed of five peer correction tasks completed by paired groups over the course of three weeks. To operationalize levels of equality and gender grouping, three composites of unequal (novice-expert) and equal (novice-novice, expert-expert) and three composites of gender grouping, including female-female, male-male (gender-matched), and male-female (gender-mixed), were formed, respectively. Finally, the learners' behaviors and reflections during and after peer correction tasks were elicited by means of observation, field notes, focused group interviews, and member-checking meetings.

Participants

Forty-three Iranian young L2 learners (22 males) aged 10 to 12 years old took part in the study. They had experienced English (L2) learning for approximately 490 hours at a non-profit language teaching institute in Kerman (Iran) prior to the data collection. Although the learners were already assigned by the administration to the pre-intermediate level of proficiency, they were not at the same levels of equality. The LoE was operationalized by classifying the subjects on the basis of their *final test scores* in the previous semester: Those achieving 87 and above in their final exam (above average) were treated as expert learners, while those whose scores fell between 75 and 86 (below average) were considered as novices (The dropping score in the test was 74 and the maximum was 100). Instructed by the same teacher, all the five pre-intermediate classrooms of the institute were selected with 7-12 learners present per class. The language for the entire activities in the institute was English, while the social language was Farsi. Therefore, the participants shared the same ethnicity, L1 background, proficiency level, L2 learning experience, instructor, and amount of L2 exposure over the course of this study.

Three textbook series were instructed as course materials: *Family Series 1 and 2* for pre-elementary level, *Hip Hip Hooray 1 to 3* for beginning level, *Hip Hip Hooray 4 and 5* for elementary level, *Touchstone 1 and 2* for pre-intermediate level, and 3 and 4 for intermediate level. Each intact classroom served as a separate treatment group (four experimental groups and one control): The experimental groups were further divided based on (a) levels of equality into equal (novice-novice and expert-expert) and unequal (novice-expert) dyads, and (b) gender dispersal into mixed (male-female) and matched (male-male and female-female) pairs. Table 1 summarizes the groups' composition in terms of levels of equality and gender dispersal.

Table 1. Group Composition in Terms of Levels of Equality and Gender Dispersal

Groups	Male	Female	Total number
EG # 1: gender-mixed + equal	3	4	7
EG # 2: gender- matched + equal	10	0	10
EG# 3: gender-mixed + unequal	2	5	7
EG # 4: gender-matched + unequal	8	0	8
CG: regular class (mixed in gender and equality)	7	4	11

Notes: EG = experimental group; CG = control group

Instrumentation

Target Items Selection

A *synonym generation test* (SGT) with 24 target words was prepared, in which the learners were required to supply the L1 equivalent of each word. Two criteria guided the selection of the target words: 1) The words were all from the new textbook (*Touchstone 1*) and were therefore unknown to the learners, and 2) The words were taken from the tasks to be performed in pairs. Overall, 18 nouns, 3 adjectives, and 3 verbs were embedded in this test.

A *grammar judgment test* (GJT) with 24 items was also devised, in which the learners were supposed to (a) differentiate between grammatically correct (by putting the letter C) and grammatically incorrect (by putting the letter I) utterances, and (b) write the correct forms next to each erroneous item. Two criteria were followed in choosing the grammatical aspects: (1) All the target features were from the learners' (previous or current) textbooks; (2) The grammatical features had a high frequency of occurrence in both textbooks (e.g., tense, number agreement, word order). Overall, the items in the GJT addressed tense, word order, subject-verb agreement, infinitive, and modal verbs. Each item carried an equal one-point mark.

The pretest as well as the immediate and delayed post-tests consisted of three versions of the synonym generation and grammar judgment tests. In the case of the synonym generation test, all the target 24 words were iterated exactly in all three tests, but grammar knowledge was tested via three parallel (structurally similar but thematically different) GJTs. Two expert reviewers (both PhD holders of Applied Linguistics) critically evaluated and commented on the tests (the three GJTs and one SGT). As a result of their reviews, several minor modifications were made to the tests. Once the tests were piloted with a small sample of learners in the same institute and of the same age and proficiency groups, the reliability indices were computed using the Cronbach alpha formula, the results of which are as follows: $\alpha_{\text{pretest (GJT)}} = .68$, $\alpha_{\text{pretest (SGT)}} = .71$, $\alpha_{\text{post-test I}} = .70$, $\alpha_{\text{post-test II}} = .69$.

Peer Correction Tasks

Learners in the treatment groups completed five tasks in pairs as extra activities. Firstly, the dyads were supposed to identify the grammatical errors in a short passage from their textbook (*Hip Hip Hooray 5*). The passages were selected from the learners' textbook since the content and meaning were familiar to the learners and the level of difficulty matched their ability level. For each task, the instructor (and the researcher as well) derived some related sentences from the textbook, put them together as a short meaningful passage (nearly six lines long), and made them ungrammatical.

Secondly, the treatment groups did a fill-in-the-blank task by selecting the appropriate words from a set of given vocabulary options. Over the course of task completion, the pairs had to cooperate with and support each other, and their partnership was the key to their success in error correction and fill-in-the-blank tasks. In the meantime, the learners' reactions and behaviors were also observed and recorded (see below for the description of observation and field notes).

Focused Group Discussion

A focused group discussion was designed to elicit the learners' perceptions and experiences toward the five peer correction tasks from multiple facets. The discussion, which was conducted in learners' first language (Farsi), consisted of 14 preplanned prompts around five main areas: (a) perceived level of motivation, anxiety, and confidence during task completion, (b) ideas about the effect of gender dispersal, degree of intimacy, nature of the tasks, and class characterization on their willingness to cooperation and tasks performance, (c) tendency to

continue these tasks in the long run, (d) attitudes towards receiving teacher's assistance, and finally (e) sources of and rationales for their tendencies and responses (see Table 2).

Table 2. Areas Covered in the Focused Group Discussion

Main parts	Orientations
Individual differences	motivation, self-confidence, anxiety
Social factors	gender, degree of intimacy (friendship)
Tasks-related factors	nature, length, level of difficulty, time limitation
Teacher role	teachers' assistance (direct or indirect)
Class-related factors	class size, atmosphere, educational system, time of the class, class activities, and class entertainment

Observation and Field Notes

Learners' actual performances, behaviors, types of interaction, and manners of feedback exchange during peer activities were observed, recorded, and later transcribed. On average, some 15 excerpts of dyadic discourses were recorded, five of which will be discussed in this research.

Member-Checking Meetings

After the study was completed, several member-checking meetings were scheduled with some purposively selected learners and held in the form of individualized interviews in Farsi. The objectives were to (a) obtain more detailed information from individual learners about their DC experiences and suggestions, (b) detect authentically the potential impact of emic perspectives, namely social and affective differences, on their DC process and outcome, and (c) buttress the data derived from the focused group discussion, observation, and field notes (Heigham & Croker, 2009).

Data Collection Procedure

Initially, the respective authorities at the institute sought permission, and the learners gave informed consent to participate in the experiment. They were ensured that their anonymity would be protected and were given advance notice that they would be doing five tasks and three test packages. Moreover, they were advised that their partnership, degree of involvement, and grades would contribute to their final course grades (i.e., inducement factor). The data were collected over five phases for 10 weeks. Table 3 outlines the timeframe of the data collection procedure for the three dyads.

Table 3. The Procedure of Data Collection for the Three Dyads

Phases	Weeks	Activities
1	1	Group assignment and pretest
	2	Peer correction tasks 1-2 + Observation and field notes
2	3	Peer correction tasks 3-4 + Observation and field notes
	4	Peer correction task 5 + Observation and field notes
3	5	Immediate post-test
4	7	Delayed post-test
5	7	Focused group discussion and member-checking meetings

After the classification, the pretests (grammar judgment and fill-in-the-blank tasks) were administered in the first session to all the participating groups to assess their initial morpho-syntactic knowledge (*t = 45 min*).

The five peer correction tasks, as a form of DC, were then designed and implemented as treatment in order to (a) determine how much the pairs can do together without any external help for solving linguistic problems and (b) form a favorable occasion for a researcher or teacher to observe and monitor learners' socio-cognitive reactions and interactions (Ellis, 2009). The treatment was offered in the experimental groups (novice-novice, expert-expert, and novice-expert in both symmetrical and asymmetrical gender pairs) over three weeks. At the beginning of each session, all pairs were engaged in a warm-up activity for five minutes, where the tutor also described the upcoming peer correction tasks and provided the rubric and guidelines in L1. Each dyad was then jointly engaged in the task, having been notified in advance that they would not have access to any aids or supplementary materials (e.g., print or online dictionaries and the *Touchstone* book) or to any scaffold on the part of the teacher, except for their partners. It took nearly 25 minutes for the pairs to complete each task, after which the papers were collected and scored by the teacher.

In order to control for memory factor as a source of test bias (Mackey & Gass, 2016), the learners were granted a one-week interval to study all the materials and prepare for the post-tests. After a week, they sat for the first GJT and SGT. The tests were carried out over 45 minutes and served as indices of the learners' morpho-syntactic intake or short-term learning. Two weeks later, the delayed post-tests, containing a parallel GJT and the SGT, were administered. The tests, which represented the learners' morpho-syntactic uptake or long-term retention, were completed over the course of 45 minutes.

Two divided focused group sessions were ordained in the same week for the experimental groups, each lasting 35 minutes. Subsequently, 11 learners from the experimental groups were individually interviewed to gain a more in-depth understanding of individual cases of particular interests. The interviews were audio-recorded in situ and common patterns of responses were later extracted. The themes extracted from the group discussion, observation, and member-checking meetings were carefully examined by all the researchers, and areas of disagreement and ambiguity were closely discussed until full consensus was achieved.

Findings

Quantitative Results: GJTs and SGTs

The Kolmogorov-Smirnov normality test was utilized to determine whether or not the data were normally distributed. Since the distribution of the data for all group compositions violated the assumption of normality ($p < .05$), nonparametric statistics was used for data analysis. Mann-Whitney U and Friedman Tests were conducted for between- and within-group comparisons (Questions 1-4), respectively. They were aimed to determine the efficacy of DC on the learners' structure and vocabulary development across the three testing conditions.

Efficacy of DC

Table 4 represents the between-group results of the treatment and control groups. As the table suggests, there are significant differences between the groups in terms of performance on the three structure tests. That is to say, the DC group, regardless of LoE and gender dispersals, has significantly outperformed the contrast group in both the immediate ($z = -3.40, p < .01$) and delayed ($z = -3.26, p < .01$) tests, implying that dyadic interaction has had both short-term and long-term retention effects.

Table 4. Intergroup Differences between DC and Control Groups (n = 43, df = 1)

Groups	Structure			Vocabulary		
	Pretest	Post 1	Post 2	Pretest	Post 1	Post 2
DC (n = 32)	23.59	25.81	25.66	22.91	25.16	24.98
Control (n = 11)	17.36	10.91	11.36	19.36	12.82	13.32
Z	-1.42	-3.40	-3.26	-.81	-2.81	-2.66
sig.	.154	.001	.001	.418	.005	.008

Within-group comparisons were also conducted for each target domain (Table 5). Unlike the treatment group which has significantly progressed in both structure ($z = 25.68$, $p < .01$, $r = .5$) and vocabulary ($z = 20.84$, $p < .01$, $r = .4$) areas, the control group has failed to achieve a significant gain in the target domains, namely structure ($z = 2.39$, $p > .05$) and vocabulary ($z = 2.60$, $p > .05$).

Table 5. Within-Group Comparisons of DC and Control Groups (n = 43, df = 2).

Groups	Structure					Vocabulary				
	Pretest	Post1	Post2	z	sig.	Pretest	Post1	Post2	z	sig.
DC	1.13	2.93	1.93	25.68	.000	1.51	2.45	2.03	20.84	.000
Control	2.32	2.00	1.68	2.39	.303	2.27	2.09	1.64	2.60	.273

According to Topping (2010), a limitation of the previous research on peer collaboration and practice is the "unbalanced group sizes" (p. 340). In Van Gennip et al. (2010), for instance, not only are the control and experimental groups not large enough, but they are widely disparate (with 28 differences) in size. Topping suggests that "future studies should use random assignment with larger group size, or alternatively matched groups with random allocation to experimental and control groups" (p. 340).

In order to address this limitation and control for the confounding effect of sample size difference, 15 cases were randomly selected from the experimental group data set. By so doing, two balanced sets of data related to the control ($n = 11$) and experimental ($n = 15$) groups were obtained.

Table 6. Intergroup Differences after Randomization (n = 26, df = 1)

Groups	Structure			Vocabulary		
	Pretest	Post 1	Post 2	Pretest	Post 1	Post 2
DC (n = 15)	14.87	17.47	17.40	14.47	16.93	16.53
Control (n = 11)	11.64	8.09	8.18	12.18	8.82	9.36
Z	1.14	9.59	9.26	.57	7.17	5.62
sig.	.285	.002	.002	.450	.007	.018

Table 6 illustrates between-group results for the two groups across the three testing conditions. Even after substantial data reduction, there is a significant difference between the experimental and control groups in both target domains. In terms of structure, the treatment

group has outperformed in the immediate ($z = 9.59, p < .01, r = 1.9$) and delayed ($z = 9.26, p < .01, r = 1.8$) tests. With respect to vocabulary too, the treatment has been to the advantage of the DC group in the first ($z = 7.17, p < .01, r = 1.4$) and second ($z = 5.62, p < .05, r = 1.1$) post-tests.

Within-group differences were recomputed after data pruning for each target domain (Table 7). Still with data reduction, the DC group performed more successfully in the immediate and delayed post-tests of structure ($z = 25.68, p < .01, r = 5.04$) and vocabulary ($z = 21.46, p < .01, r = 4.2$) compared with their pretests performance. Between- and within-group comparisons conducted after randomization provide further evidence that the DC group has had more language gains than the non-collaborative group, and that dyadic collaboration has been substantially efficacious in L2 development.

Table 7. Within-Group Comparisons after Randomization (n = 26, df = 2)

Groups	Structure					Vocabulary				
	Pretest	Post 1	Post 2	z	sig.	Pretest	Post 1	Post 2	z	sig.
DC	1.13	2.93	1.93	25.68	.000	1.17	2.80	2.03	21.46	.000
Control	2.32	2.00	1.68	2.39	.303	2.27	2.09	1.64	2.60	.273

LoE as Mediator I. Table 8 illustrates the results for the performance of the two equal groups on structure and vocabulary tests.

Table 8. Within-Group Comparisons of Equal Dyads (df = 2)

Groups	Structure					Vocabulary				
	Pre	Post 1	Post 2	z	sig.	Pre	Post 1	Post 2	z	sig.
Expert (n=8)	18.50	21.00	20.00	13.06	.001	20.50	23.00	20.00	8.07	.018
Novice (n=8)	8.50	9.50	9.50	3.12	.210	10.00	14.00	10.50	4.83	.089

Notes: Pre: pretest, Post 1: immediate post-test, Post 2: delayed post-test

The expert-expert group has substantially improved in both structure ($z = 13.06, p < .01$) and vocabulary ($z = 8.07, p < .05$) areas. That is to say, structure and vocabulary, both with large effect sizes of 2.67 and 1.65, respectively according to Cohen's (1988) criterion, account for a significant variance in the expert-expert group scores. On the contrary, within-group comparison of novice-novice dyads demonstrates no significant differences in structure ($z = 3.12, p > .05$) and vocabulary ($z = 4.83, p > .05$) across the three

testing conditions. To sum up, peer interaction in equal groups has been more effective for an expert than novice dyads.

Table 9 summarizes the results for unequal groups in terms of performance on structure and vocabulary tests. Although unequal groups have significantly outperformed in structure (*z=19.27, p<.01*) throughout the course, they have failed to experience much gain in vocabulary (*z=5.73, p>.05*).

Table 9. Within-Group Comparisons of Unequal Dyads (df=2)

Composition	Structure					Vocabulary				
	Pre	Post 1	Post 2	z	sig.	Pre	Post 1	Post 2	z	sig.
Expert-Novice	15.00	18.50	17.50	19.27	.000	15.00	18.00	17.00	5.73	.057

Notes: Pre: pretest, Post 1: immediate post-test, Post 2: delayed post-test

For this reason, expert and novice learners in unequal groups were investigated separately from each other in terms of structure to find out which group has benefitted more from peer feedback activities (Table 10).

Table 10. Comparisons of Experts' and Novices' Structure Gains in Unequal Dyads (df=2)

Groups	Structure				
	Pre	Post 1	Post 2	z	sig.
Expert (n = 8)	17.50	22.50	21.00	12.28	.002
Novice (n = 8)	8.00	17.00	15.00	15.54	.000

Notes: Pre: pretest, Post 1: immediate post-test, Post 2: delayed post-test

According to Table 10, both groups have significantly improved in grammar learning as a result of involvement in peer activities in unequal dyads. It implies that in unequal group composition, structure learning, with a large effect size for both expert (*r=2.51*) and novice (*r=3.17*) learners, has substantially explained their variations in the three GJTs scores.

Gender Grouping as Mediator II. Table 11 summarizes the performance of gender-equal dyads across structure and vocabulary pretests and post-tests.

Table 11. Within-Group Comparisons of Gender-Matched Dyads (df = 2)

Groups	Structure					Vocabulary				
	Pre	Post1	Post2	z	sig.	Pre	Post1	Post2	z	sig.
Female (n = 8)	10.50	20.50	18.00	15.54	.000	10.00	18.50	14.00	12.60	.002
Male (n = 8)	12.50	15.00	14.00	12.48	.002	14.00	19.00	16.00	11.46	.003

Notes: Pre: pretest, Post 1: immediate post-test, Post 2: delayed post-test

Female-specific groups have experienced significant progress in structure ($z = 15.54$, $p < .01$, $r = 3.17$) and vocabulary ($z = 12.60$, $p < .01$, $r = 2.57$) domains. Male participants too have achieved significantly higher scores on structure ($z = 12.48$, $p < .01$, $r = 2.55$) and vocabulary ($z = 11.46$, $p < .01$, $r = 2.34$) post-tests compared with the pretests. Conclusions can be drawn on the basis of the results reported in Table 11 that male and female learners, when participating in gender-matched paired activities, have improved equally well in L2 grammar and vocabulary development.

Table 12. Within-Group Comparisons of Gender-Unmatched Dyads (df = 2)

Composition	Structure					Vocabulary				
	Pre	Post 1	Post 2	z	sig.	Pre	Post 1	Post 2	z	sig.
Female-Male	15.00	18.50	17.50	19.27	.000	15.00	18.00	17.00	5.73	.057

Notes: Pre: pretest, Post 1: immediate post-test, Post 2: delayed post-test

Finally, the results of the gender-unmatched dyads' performance on structure and vocabulary tests are depicted in Table 12. The groups, which were mixed in gender, have performed poorly across the three tests of structure ($z = 3.25$, $p > .05$) and vocabulary ($z = 2.47$, $p > .05$), with no statistically significant enhancement over the course of instruction. This finding suggests that gender-unmatched paired grouping has relatively been to the disadvantage of both male and female partners. It is likely that gender differences have hindered them from interacting efficiently enough to facilitate and smooth language learning process. Table 13 supplies a summary of the quantitative results.

Table 13. Summary of the Quantitative Results

Groups/Dyads	Structure		Vocabulary	
	STL	LTR	STL	LTR
DC group > non-DC group	√	√	√	√
DC group	√	√	√	√
Non-DC group	n.s.	n.s.	n.s.	n.s.
Expert-expert dyad	√	√	√	√
Novice-novice dyad	n.s.	n.s.	n.s.	n.s.
Experts in expert-novice dyad	√	√	n.s.	n.s.
Novices in expert-novice dyad	√	√	n.s.	n.s.
Male-male dyad	√	√	√	√
Female-female dyad	√	√	√	√
Males in male-female dyad	n.s.	n.s.	n.s.	n.s.
Females in male-female dyad	n.s.	n.s.	n.s.	n.s.

Notes: DC = dyadic collaboration, STL = short-term learning, LTR = long-term retention

Qualitative Results: Observation and Interviews

Observation and Field Notes

The following excerpts from observational data illustrated different patterns of dyadic interaction, manners of feedback, and the interlocutors' behaviors and reactions. Five dyadic discourses are reproduced below.

Excerpt # 1. Two participants in an equal (E1-E2) and gender-matched (both females) pair contributed jointly to the fulfillment of a grammatical judgment task. This dyadic discourse contains numerous instances of confirmation, clarification request, hypothesis testing, non-directive implicit feedback, and metalinguistic talk. It actually indicates that the two experts successfully reached resolutions via a process of meaningful negotiation.

E1: Will travelled…next week…what is it?

E2: Mm…future tense?

E1: Yeah, I think so …mm…/will/ shows activities in future.

E2: Uh, you're right…

E1: /Next week/… time expression for future tense.

E2: Yes… travelled? /ed/ is for the past tense… will travelling?

E1: /ed/ for the past? Yeah…mm…will travelling? Why?

E2: Uh… we use /v-ing/ for the past, now and future?

E1: Yeah…but… it is present continuous…mm…this pattern… am, is, are + verb-ing

E2: Verb /to be/ + v-ing?

Excerpt # 2. A dyadic interaction between two experts in a mixed-gender pair is illustrated below. Although both partners were experts and engaged, it was not a joint contribution. This dyadic discourse is characterized by considerably fewer uses of clarification requests and positive feedback (i.e. confirmation). Rather, it consists of a substantial share of disagreement, resistance to accept each other's comments, and insistence on own opinions (e.g., reflected in the use of the first singular pronoun). It was also observed that this dyad often communicated in a raised voice and irritated verbal tone and experienced frustration and resentment.

F: That uh...she was in the library?

M: Ah

F: In a library

M: No no no no …. in the library.… not in a library...mm... the problem is not /the/

F: Okay I myself know it.

M: I think...on the library

F: No no no... /on/ is not a preposition of place

M: No no …. Listen...on the table.

F: Think... table is not place...

M: Mm... at the library...

F: At the... no no no

M: I will put at...

Excerpt # 3. The third excerpt was extracted from the data of two participants in an N-E and matched-gender (female) pair. Here, the expert sought to involve the novice in the interaction and provide assistance and explanations. Some features of this dyadic discourse were prompts such as encouragement, questions, confirmations and positive feedback, implicit feedback, and reformulations mostly on the part of the expert partner.

E: She can dancing and whistling?

N: Mm... she?...can dances?...whistles?

E: Pay attention... we have /can/...

N: Yeah...uh...can?

E: Can I go? Can I come in?

N: Huh...can go? Can come?

E: Exactly.… So?

N: Yeah...mm...can dance? ...whistle?

E: Excellent... can read... can walk... can say...

N: Huh no /ed/ no /ing/ no /-s/... can run... can write...

E: Yeah... good.

E: She can dance and whistle.

Excerpt # 4. Excerpt 4 presents an interactional episode between two novices jointly involved in a grammatical judgment task. This interaction contains unclear comments, hesitations, pauses, and incorrect suggestions to the point of representing a vague and misguided communication.

N1: I would like listening...?

N2: Okay...mm...

N1: Think about it....

N2: Well

N1: I liked...

N2: I liked to listening?

N1: I.... I guess...

N2: Mm...maybe

N1: What do you think?

N2: I don't know ... mm...no idea

N1: Okay... Put /I liked/...

Excerpt # 5. In Excerpt 5, two instances of interaction between an N-E dyad (both males) over a fill-in-the-blank (vocabulary) task are unveiled. Both dyadic discourses are marked by direct and explicit feedback, minimal negotiation, and quick resolutions.

E: Okay..... Which word?

N: Mm... /ferry/?

E: No...

N: Yeah..... mm....What is the meaning of /ferry/?

E: I don't know.

N: Let's go to the next one.... We will get back...

E: I think, the answer is /sneakers/.

N: Mm... Okay. What about this one?

E: Mm .../put on/.

N: The meaning of put on?

E: Wear?... Mm... I am not sure...

N: Okay, I will check it at home...

Focused Group Discussion

A high level of perceived effectiveness of DC was recorded in N-E and E-E as well as mixed-gender and matched-gender pairs (see E-file for a selected sample of learners' direct quotations). On the contrary, N-N pairs contended that DC was not practical due to various reasons. The participants confirmed that the intimacy, friendship, and responsibility shared during DC made them more attentive and active, decreased their stress, and improved their micro-skills. Besides, 31 out of the 34 learners (91.17%) pointed out that paired activities were fun and enjoyable.

Overall, the learners were willing to continue peer correction tasks in the long run, while all the learners in N-N pairs preferred to pick up a partner on their own. All the male learners preferred female partners. In this respect, females also favored acting in gender-matched pairs, while a few of them pointed out that gender is not a determining causal factor.

Respecting self-confidence and motivation in feedback generation, some learners, in particular females, pointed out that they had enough confidence in themselves. Twenty-seven (79.40%) learners, particularly novices in N-E pairs, reported an increased self-confidence after DC activities. Experts in N-E pairs, however, complained that their partners refused to offer feedback due to poor self-assurance. Moreover, males, regardless of their levels of equality, believed that the additional score was their only source of motivation (extrinsic motivation). On the contrary, females commented that they were initially driven by the extra credit but gradually found dyadic tasks more encouraging and helpful than solo activities for language learning (intrinsic motivation).

When enquired about the role of teacher's assistance, learners in N-E and E-E pairs preferred to receive teacher feedback after task completion, while learners in N-N pairs demanded it during the tasks, probably as a result of the limited knowledge shared between them. Regarding the impact of learners' levels of equality on the efficacy of DC, most learners suggested that being committed and cooperative serves a more substantial role than the partner's levels of equality does. Moreover, for 30 out of 34 (88.25%) the fill-in-the-blank (vocabulary) tasks were more tedious than grammar tasks.

Finally, the last interview question centralized on task-related factors including time, length, and level of difficulty. All the participants stated that the time and length of the tasks were fine-tuned, while fill-in-the-blank tasks were more difficult to complete than grammatical judgment tasks.

Member-Checking Meetings

Six major themes emerged from member checking meetings, which are discussed below (see E-file for a selected sample of learners' direct quotations).

Overall preference to work with females. The first recurrent theme was the interviewees' preference for collaboration with females. The most salient reason was female learners' higher sense of responsibility and greater commitment to task completion.

Experts' tendency to collaborate with novices. A number of expert learners believed that working with novices was easier and more enjoyable.

Experts' tendency to work individually. Some experts notified that DC made them bored and exhausted since they could accomplish the tasks on their own more quickly and effortlessly.

Novices' inclination to pair up with experts. All the novices in N-E pairs were positive towards their interactions and found their partners' feedback as helpful and reliable as the teacher's.

DC as more effective for grammar than vocabulary. Learners maintained that they enjoyed implementing grammatical judgment tasks more than fill-in-the-blanks. They reasoned that vocabulary activities were more difficult as a result of their unsystematic form-meaning associations and were less salient considering the course objectives and exam requirements.

Personality factors as determinants of DC quantity and quality. Although both experts and novices asserted that pooling feedback was the key to success in peer activities, a few of them pointed out that shyness, uncertainty, and self-doubt prevented them from making enough contributions.

Discussion

The first two research questions concerned the short- and long-term effects of DC on young language learners' morpho-syntactic development. The results were positive. In the course of collaborative activities, pairs generated and received a considerable amount of feedback,

which is fairly inconceivable on the part of teachers considering their heavy schedules and workloads (Cho et al., 2006; Gielen et al., 2010). Based on the observation notes, the tutor in the control group could only sporadically respond to the learners' errors, while learners exchanged abundant and diverse feedbacks in the process of peer correction. The results are also evidenced in other EFL settings (e.g., Kowal & Swain, 1994; McDonough, 2004; Shehadeh, 2011; Storch, 2007), all of which suggest that learners' active involvement in advisements over language helps promote performance in both grammar and vocabulary areas.

To answer the third question, learners' overall accuracy was compared across three variations of LoE. The results suggested that, unlike the novice-novice dyads, expert-expert pairs attained significant improvement over the course of collaborative instruction. This finding is not surprising given that the novices in equal pairs shared markedly fewer corrective feedbacks as a result of having less extensive, elaborated, and domain-peculiar knowledge (Lundstrom & Baker, 2009). Moreover, chances are high that novices share ambiguous, incorrect, and misleading comments due to their limited knowledge, which may bear even adverse effects on their language acquisition (Cho & MacArthur, 2010; Niu et al., 2018). This possibility was reinforced by the observation results since novices' communication was characterized by loosely organized comments, confusing feedbacks, long look-ups and pauses, and hesitations. In contrast with the present study, all the learners in Watanabe (2008) improved through paired problem-solving tasks regardless of their levels.

On the other hand, expert-expert pairs' progress can be accounted for by the concept of *intersubjectivity*. The concept refers to the mutual understanding that is achieved between people through effective communication (Fosnot & Perry, 2005). Noticeable in both quantitative and observation results were mutual understanding, effective communication, and meaningful interaction between experts due to the co-availability of language knowledge and problem-solving skills. In fact, experts are better capable of generating comprehensible feedback, synthesizing individual thoughts, and negotiating actively to solve language problems, particularly with an ability-matched partner. Communication between two expert partners entails fewer clues, faster problem detection, automatic responses, and more efficient usage of heuristic searches and rules, which enables them to accomplish tasks in a more systematic and effective manner (Besnard & Bastien-Toniazzo, 1999).

Another LoE-related finding was that both experts and novices benefited equally well from unequal pairing, but in the grammar sphere only. It gives credence to the results of Lundstrom and Baker (2009) which found that the peer review process is to the benefit of

givers (assessors) and receivers (assessees) alike. In a similar vein, Allen and Mills' (2014) study indicated when novice learners collaborate with expert peers, they have the opportunity to receive rich comments and diverse examples as feedback. Our finding, however, is inconsistent with Williams' (2001) study which concluded that novice learners may not necessarily benefit from collaborative tasks with respect to their language accuracy. The effectiveness of DC for expert partners in unequal pairs supports the results of Lundstrom and Baker (2009), Min (2006), and Rollinson (2005). According to these studies, learning to systematically review peers' productions and attending to their knowledge gaps may ultimately make better self-reviewers of the assessors, enabling them to critically monitor and evaluate their own performance and revise it. According to Holunga (1994), "verbalizing the language errors that the learners spotted allowed them to become aware of their problems, set goals for themselves, monitor their own language use, and evaluate their overall success" (p. 109).

Relevantly, another intriguing finding was the contribution of DC in unequal pairs to structure but not vocabulary development. Several explanations are conceivable. Firstly, the improved accuracy in grammar may be related to the longer time and interaction dedicated to grammar judgment tasks. Based on the observation notes, the time taken to complete the grammatical tasks in pairs almost doubled the time taken to complete the synonym generation tasks. In addition, it was documented in the observation and focused group discussion phases that learners were more interested in fulfilling grammatical tasks than vocabulary items, which is most likely driven by the prevailing structure-centered orientation of the educational setting under investigation. A similar argument was offered by Storch (1999) when her learners were found more motivated to concentrate on grammatical accuracy.

An alternative explanation is the *involvement load hypothesis*. Based on the theory, the greater the learners *need* an item and the longer it takes them to *search* and *evaluate* possible alternatives, the more efficiently the item is learned and retained. This finding, in compliance with the previous analyses (e.g., Cho & MacArthur, 2010; Shehadeh, 2011), suggests that active involvement with negative evidence (feedback) followed by modified self- or peer-initiated output improves productions. In the present study, negative evidence and modified output were observed in grammar judgment tasks only.

Thirdly, but no less importantly, is the *rule-governed* nature of grammar. In sharp opposition to vocabulary, learning grammar is guided by a set of rules and provokes a systematic and formulaic approach. This makes it much easier and more practical for learners to acquire and recall grammatical items through instructed language learning.

Further analyses also divulged that both males and females were significantly more

successful when engaged in gender-matched dyadic interactions than in mixed pairs. The finding can be rationalized on two grounds, namely gender-related differences and cultural exigencies of the educational context under study. Empirical research has characterized females as more outspoken than males, more willing to correct peers, inclined towards rapport (interactional) talk to a certain other in pair, and in favor of interacting in an affective, flexible, and caring manner (Pishghadam & Kermanshahi, 2011; Tannen, 1990). Observation notes also made female partners known as more explicit and systematic in commenting than males, often providing feedbacks in the form of patterns and formulas (metatalk), whereas males supplied implicit feedback using down-toning devices or fleeting comments. Inherent gender incongruences of these types are, at least in part, responsible for the ineffective collaboration in mixed-gender pairs and their failure to reach the expected language benefits. In other words, the proportion of mutual understanding and meaningful negotiation, as the major qualities of DC and determinants of language success, has been considerably larger in gender-equal pairs (Strijbos et al., 2010). The sociocultural background of the learners could also be at play. Iranian learners do not experience mixed-gender groupings until tertiary-level education, which could normally result in disputes, misinterpretations, and underachievement among gender-unmatched dyads. Therefore, as stated by Pawlak (2014), educational decisions, here pairing up the learners, must reflect "educational traditions, curricular requirements, [learners'] deeply ingrained beliefs … and prevalent expectations" (p. 84).

Finally, the learners' behaviors and reactions during and following the DC activities were addressed (Research Question 5). The first items addressed their views of the (dis)advantages of DC across levels of equality. Group discussion and member checking meetings revealed that the majority of learners (70.60%), both novice and expert, were in favor of paired activities and found them practical in multiple ways. This finding supports the previous studies which regard awareness raising and gap noticing as two practical products of paired tasks (e.g., Dobao, 2012; Kim, 2008; Storch, 2005; Van Gennip et al., 2010; Wiggleswoorth & Storch, 2009). However, learners in novice-novice pairs were less positive towards DC and reported fewer perceived language gains due to their equally poor knowledge, which corroborates the findings of Allen and Mills (2014), Cho and MacArthur (2010), and Lundstrom and Baker (2009).

In addition, the majority of the learners (91.17%) found peer correction tasks enjoyable and fun-filled. Most of them pointed out that in the early stages, the tasks were confusing and intimidating, but after a while mistakes, in particular the absurd ones, created lots of joy and laughter and enabled them to offer their ideas and comments without feeling embarrassed any

longer. They particularly referred to rapport and friendship as the major affective outcomes of DC. It must be mentioned here that the sample knew each other well since they were classmates for around four years, but they reported a higher level of intimacy after these dyadic tasks. This lends evidence to the previous studies which suggested that learners enjoy paired experiences as a source of learning (e.g., Shehadeh, 2011; Storch, 2005).

Thirdly, more than two-thirds of the learners (particularly novices in N-E pairs) disclosed higher self-confidence after the DC experience. One possible explanation could be the tendency of the experts to underestimate the difficulty level of the tasks, inducing the impression that the novices were able to perform the tasks as effectively as the experts did. Thus, this underestimation of difficulty boosted their self-confidence. This finding is in contrast with that of Camerer et al. (1989) which introduced the *curse of expertise* as a common expert behavior, prompting the highly proficient to underestimate the difficulty level of tasks and therefore to be ineffective in transmitting skills and knowledge to novices. It also contradicts Storch's (2007) study in which the participants felt less confident and relaxed over the paired activities as a result of being corrected and criticized on the peers' part.

Learners' level of anxiety and trust in the assessor's knowledge were the next factors explored in this study. All the learners, irrespective of their LoE and gender grouping, asserted that they experienced little stress during dyadic tasks due to high levels of intimacy and empathy. Moreover, in contrast to the N-N pairs who had learned more from teacher feedback (previously practiced) than from peers (in the current semester), the novices in N-E pairs reported perfect trust in the experts' knowledge and divulged that their expert partners were as resourceful as the teacher. This association of the trust issue with LoE was actualized well in the statistical results in that the learners in N-E and E-E pairs, regardless of their gender, outperformed N-N pairs and the control group.

Another concern was if the learners were in favor of paired activities in the long run. The participants asserted that they are, provided that they have a desirable partner or can select their partner on their own. The interest was particularly highest among novices in unequal pairs since they believed that systematic collaboration with a more knowledgeable peer could markedly contribute to their educational growth. Similar findings were reached by Storch (2005), Shehadeh (2011), and Van Gennip et al. (2010).

Finally, in response to the question of whether they preferred a male or female partner, both genders favored females. They asserted that females were more cooperative and committed than males, generated more comments on their own initiative, and were more supportive and caring particularly with less proficient partners. This finding is in contrast

with the studies that concluded that the proportion of shared feedback and negotiations was larger in gender-equal than cross-gender pairs (e.g., Strijbos et al., 2010).

Conclusions and Implications

The research addressed DC in relation to two variables of levels of equality and gender grouping in a typical South Asian EFL context. Albeit contrasting in a few aspects under study, the results of the quantitative (test performance) and qualitative (observation and interviews) phases matched in most areas. While asymmetry relatively worked for the LoE factor, it was not necessarily effective for gender grouping since DC was helpful for expert-expert dyads, gender-matched pairs, both experts and novices in unequal pairs, but was ineffective for novice-novice and cross-gender dyads. Moreover, the learners, regardless of their gender, favored to pair up with females and partners of opposite LoE.

From a pedagogical standpoint, then, language teachers are recommended to take into account the learners' preferences and perceptions in educational decision-making. Dyadic tasks can also be implemented as part of language classroom activities, but it is suggested that the sample size be larger and learners be paired up with partners of their own choice. As this study contained N-N in matched-gender pairs and E-E in mixed-gender pairs only, future research is invited to reserve N-N dyads in mixed-gender and E-E in matched-gender dyads for more valid comparisons.

References

Allen, D., & Mills, A. (2014). The impact of second language proficiency in dyadic peer feedback. *Language Teaching Research*, *20*(4), 1-16. Doi: 10.1177/1362168814561902

Aries, E. (1976). Interaction patterns and themes of male, female, and mixed groups. *Small Group Behavior*, *7*(1), 7-18.

Besnard, D., & Bastien-Toniazzo, M. (1999). Expert error in trouble-shooting: An exploratory study in electronics. *International Journal of Human Computer Studies*, *50*(5), 391-405.

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic setting: An experimental analysis. *The Journal of Political Economy*, *97*(5), 1232-1254.

Caspi, A., & Blau, I. (2011). Collaboration and psychological ownership: How does the tension between the two influences perceived learning?. *Social Psychology of Education*, *14*(1), 283-298.

Chang, L. (2007). The influence of groups processes on learners' autonomous beliefs and behaviors. *System*, *35*(3), 322-337.

Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, *20*(4), 328-338.

Cho, K., Schunn, C., & Wilson, R. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, *98*(4), 891-901.

Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences.* Lawrence Erlbaum Associates.

Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2018). What makes for effective feedback: Staff and student perspectives. *Assessment & Evaluation in Higher Education*, *44*(1), 25–36. Doi: 10.1080/02602938. 2018.1467877

Diab, N. (2010). Effects of peer-versus self-editing on students' revision of language errors in revised draft. *System*, *38*(1), 85-95.

Dobao, F. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work. *Journal of Second Language Writing*, *21*(1), 40-58.

Dobao, F., & Blum, A. (2013). Collaborative writing in pairs and small groups: Learners' attitudes and perceptions. *System, 41*(2)*, 365-374.* Doi: 10.1016/j.system.2013.02.002

Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal*, *63*(2), 97-107.

Elola, I., & Oskoz, A. (2010). Collaborative writing: Fostering foreign language and writing conventions development. *Language Learning and Technology*, *14*(3), 51-71.

Fosnot, C., & Perry, R. (2005). Constructivism: A psychological theory of learning. In C. Fosnot (Eds.), *Constructivism: Theory, perspectives and practice* (pp. 8-38). New York: Teacher's College Press.

Garrett, P., & Shortall, T. (2002). Learners' evaluations of teacher-fronted and student-centered classroom activities. *Language Teaching Research*, *6*(1), 25-57.

Gass, S., & Mackey, A. (2007). Input, interaction and output in SLA. In B. van Patten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp.175-199). Mahwah, NJ: Lawrence Erlbaum.

Ghahari, S., & Farokhnia, F. (2017). Triangulation of language assessment modes: Learning benefits and socio-cognitive prospects. *Pedagogies: An International Journal*, *12*(3), 275-294. https://doi.org/10.1080/1554480X.2017.1342540

Ghahari, S., & Farokhnia, F. (2018). Peer versus teacher assessment: Implications for CAF

triad language ability and critical reflections. *International Journal of School and Educational Psychology*, *6*(2), 124-137. https://doi.org/10.1080/21683603.2016.1275991

Ghahari, S., & Piruznejad, M. (2017). Recast and explicit feedback to young language learners: Impacts on grammar uptake and willingness to communicate. *Issues in Language Teaching*, *5*(2), 209-187. Doi: 10.22054/ilt.2017.8058

Ghahari, S., & Sedaghat, M. (2018). Optimal feedback structure and interactional pattern in formative peer practices: Students' beliefs. *System*, *74*, 9-20. Doi: 10.1016/j.system.2018.02.003

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.

Heigham, J., & Croker, R. (2009). *Qualitative research in applied linguistics*. Basingstoke, UK: Palgrave Macmillan.

Holmes, J. (1992). *An introduction to sociolinguistics.* London: Routledge.

Holunga, S. (1994). *The effect of metacogntive strategy training with verbalization on the oral accuracy of adult second language learners* (Unpublished doctoral dissertation). Canada: University of Toronto.

Huisman, B., Saab, N., Van Driel, J., & Van Den Broek, P. (2019). A questionnaire to assess students' beliefs about peer-feedback. *Innovations in Education and Teaching International*, *57*(3), 328-338. https://doi.org/10.1080/14703297.2019.1630294

Kim, Y. (2008). The contribution of collaborative and individual tasks to the acquisition of L2 vocabulary. *Modern Language Journal*, *92*(1), 114-130.

Kim, Y., & McDonough, K. (2008). The effect of interlocutor proficiency on the collaborative dialogue between Korean as a second language learners. *Language Teaching Research*, *12*(2), 211-234.

Kowal, M., & Swain, M. (1994). Using collaborative language production tasks to promote students' language awareness. *Language Awareness*, *3*(2), 73-93.

Lo, J., & Hyland, F. (2007). Enhancing students' engagement and motivation in writing: The case of primary students in Hong Kong. *Journal of Second Language Writing*, *16*(4), 219-237.

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer reviewer's own writing. *Journal of Second Language Writing*, *18*(1), 30-43.

Mackey, A., & Gass, S. (2016). *Second language research: Methodology and design*. New York: Routledge.

McDonough, K. (2004). Learner-learner interaction during pair and small group activities in a Thai EFL context. *System*, *32*(2), 207-224.

Mercader, C., Ion, G., & Díaz-Vicario, A. (2020). Factors influencing students' peer feedback uptake: instructional design matters. *Assessment & Evaluation in Higher Education*, *43*(8), 1315-1325. Doi: 10.1080/02602938.2020.1726283

Min, H. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, *15*(2), 118-141.

Nassaji, H., & Tian, J. (2010). Collaborative and individual output tasks and their effects on learning English phrasal verbs. *Language Teaching Research*, *14*(4), 397-419.

Niu, R., Jiang, L., & Deng, Y. (2018). Effect of proficiency pairing on L2 learners' language learning and scaffolding in collaborative writing. *Asia-Pacific Education Researcher*, *27*(3), 187–195. Doi: 10.1007/s40299-018-0377-2

Otha, A. (1995). Applying sociocultural theory to an analysis of learner discourse: Learner-learner collaborative interaction in the zone of proximal development. *Issues in Applied Linguistics*, *6*(2), 93-121.

Pawlak, M. (2014). *Error correction in the foreign language classroom: Reconsidering the issues.* Berlin: Springer.

Pishghadam, R., & Kermanshahi, P. (2011). Peer correction among Iranian English language learners. *European Journal of Educational Studies*, *3*(2), 217-228.

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

Shehadeh, A. (2011). Effects and student perceptions of collaborative writing in L2. *Journal of Second Language Writing*, *20*(4), 286-305.

Simonsmeier, B., Peiffer, H., Flaig, M., & Schneider, M. (2020). Peer feedback improves students' academic self‑concept in higher education. *Research in Higher Education*, *61*, 706–724. Doi: 10.1007/s11162-020-09591-y

So, H., & Brush, T. (2008). Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. *Computers & Education*, *51*(1), 318-336.

Storch, N. (1999). Are two heads better than one? Pair work and grammatical accuracy. *System*, *27*(3), 363-374.

Storch, N. (2001). How collaborative is pair work? ESL tertiary composing in pairs. *Language Teaching Research*, *5*(1), 29-53.

Storch, N. (2002). Patterns of interaction in ESL pair work. *Language Learning*, *51*(1), 119-159.

Storch, N. (2005). Collaborative writing: Product, process and students' reflections. *Journal of Second Language Writing*, *14*(3)*,* 153-173.

Storch, N. (2007). Investigating the merits of pair work on a text editing task in ESL classes. *Language Teaching Research*, *11*(2), 143-159.

Storch, N. (2008). Metatalk in a pair work activity: Level of engagement and implications for language development. *Language Awareness*, *17*(2), 95-114.

Storch, N. (2011). Collaborative writing in L2 contexts: Processes, outcomes, and future directions. *Annual Review of Applied Linguistics*, *31*(1), 275-288.

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

Tannen, D. (1990). *You just don't understand: Women and men in conversation*. New York: William Morrows & Co.

Topping, K. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, *20*(4), 339-343.

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

Van Lier, L. (2014). *Interaction in language curriculum: Awareness, autonomy and authenticity*. Landon: Longman.

Watanabe, Y. (2008). Peer-peer interaction between L2 learners of different proficiency levels: Their interactions and reflections. *Canadian Modern Language Review, 64*(4), 605-636.

Watanabe, Y., & Swain, M. (2007). Effects of proficiency differences and patterns of pair interaction on second language learning: Collaborative dialogue between adult ESL learners. *Language Teaching Research*, *11*(2), 121-142.

Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity, and accuracy. *Language Testing*, *26*(3)*,* 445-466.

Williams, J. (2001). The effectiveness of spontaneous attention to form. *System*, *29*(3), 325-340.

Yu, S. (2019). Learning from giving peer feedback on postgraduate theses: Voices from Master's students in the Macau EFL context. *Assessing Writing*, *40*(1), 42-52.

Yu, S., & Lee, I. (2016). Understanding the role of learners with low English language proficiency in peer feedback of second language writing. *TESOL Quarterly*, *50*(2),

483-494. Doi: 10.1002/tesq.301

Yu, S., Zhang, Y., Zheng, Y., Yuan, K., & Zhang, L. (2018). Understanding student engagement with peer feedback on master's theses: A Macau study. *Assessment & Evaluation in Higher Education*, *44*(1), 50-65. Doi: 10.1080/02602938.2018.1467879

Yule, G., & Macdonald, D. (1990). Resolving referential conflicts in L2 interaction: The role of proficiency and interactive role. *Language Learning*, *40*(4), 539-556.