The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL

van Steendam, E.; Rijlaarsdam, G.; Sercu, L.; van den Bergh, H.

DOI
10.1016/j.learninstruc.2009.08.009

Publication date
2010

Document Version
Final published version

Published in
Learning and Instruction

Learning and Instruction 20 (2010) 316–327

Learning and Instruction

www.elsevier.com/locate/learninstruc

The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL

Elke Van Steendam [a,c,*], Gert Rijlaarsdam [b,d], Lies Sercu [c], Huub Van den Bergh [b,d]

[a] Faculty of Applied Economic Sciences, University of Antwerp, Prinsstraat 13, B-2000 Antwerp, Belgium
[b] Graduate School of Teaching and Learning, University of Amsterdam, Spinozastraat 55, 1018 HJ Amsterdam, the Netherlands
[c] Department of Language and Education, Faculty of Arts, University of Leuven, Blijde-Inkomststraat 21, B-3000 Leuven, Belgium
[d] Utrecht Institute of Linguistics, Faculty of Arts, University of Utrecht, Trans 10, 3512 JK Utrecht, the Netherlands

Abstract

Studies in peer feedback on written texts show that instruction in revision is necessary for the effectiveness of global feedback. Participants in the study were 247 university freshmen, native speakers of Dutch, who took the same Business English course, and were instructed a revision strategy following Schunk and Zimmerman's social cognitive model. Participants were first instructed through observation or practising, followed by dyadic or individual emulation, with the aim to determine the most effective combination of instruction and emulation for revision. Results showed a significant interaction of the above two factors. If emulation happens individually, then observation and practice are equally effective in terms of strategy acquisition. For dyadic emulation to be productive, it needs to be preceded by observation.

Keywords: Collaborative revision; Observational learning; Strategy instruction in revision; Collaborative learning; Peer feedback; Emulation; Modelling

1. Introduction

The benefits of peer collaboration for revision have been shown in several studies both in students' first language (L1) (Boscolo & Ascorti, 2004) and students' second (L2) or foreign language (FL) (De Guerrero & Villamil, 1994). Peer response, the practice of letting students discuss each other's written work orally or in writing (peer review), has already shown its beneficial effects on learning-to-write and learning-to-revise both for L2 and FL students (Berg, 1999; Min, 2005) as well as for L1 students (Zhu, 1995). The advantages of having students engage in revision with peers range from "an increased audience awareness" (Mendonca & Johnson, 1994) to improved "text quality" for the writer receiving feedback (Min, 2006) but also for the feedback provider (Rijlaarsdam & Couzijn, 2000a). However, one of the crucial conditions for revision with peers to be effective is the presence of instruction and training (Stanley, 1992).

1.1. Training in peer feedback

Studies on L2 feedback show that, without training, (novice) L2 revisers mainly focus on the surface level of a peer's text and rarely comment on structure and content (Leki, 1990; Mangelsdorf & Schlumberger, 1992). Leki (1990) writes that students who learn English as a foreign language (EFL), especially the ones who are new to the practice of responding to peer-writing, tend to focus on surface errors instead of "grappling with the more difficult question of meaning" (p. 9). When analysing the written comments EFL students made on each other's writing, Mangelsdorf and Schlumberger (1992) found that the majority of student comments reflected a so-called prescriptive stance in which students focused on correct form rather than on "the communication of meaning" (p. 235). Flynn (1982) claims that without training students mainly address lower-order

* Corresponding author. Tel.: +32 16 890 495; fax: +32 16 324 767.
E-mail address: elke.vansteendam@telenet.be (E. Van Steendam).

0959-4752/$ - see front matter
doi:10.1016/j.learninstruc.2009.08.009

concerns and surface errors and feel at a loss as to how to comment on a peer's draft because of lack of critical skills and knowledge about criteria.

The majority of studies on peer feedback, both oral and in writing, stress the need for training to enable students to give adequate feedback. Most highlight the importance of training students to enable them to give polite, non-judgmental feedback (cf. Tang & Thitecott, 1999). However, fewer studies interpret adequate feedback first and foremost as a focus on higher-order concerns (idea development, focus and structure) in a peer's draft and insist that students also need to be trained in applying evaluative criteria to a text and, in detecting, diagnosing and remedying these higher-order concerns (Min, 2005; Zhu, 1995).

A recent pretest/posttest comparison study by Min (2005) revealed that students extensively coached in peer reviewing generated more specific and relevant written feedback on global features of their peer's composition after training than before. Min (2005) trained 18 EFL Taiwanese sophomore students, who had English as major course, to give more adequate and specific written feedback. Training consisted of two phases. First the instructor demonstrated how to comment on a peer's draft following a 4-step strategy based on "characteristics of comments that were found to facilitate students' revisions" (Min, 2005, p. 293) according to prior research (Stanley, 1992; Tang & Thitecott, 1999). Students were encouraged to always take the following four steps into account when commenting on a peer's draft: (a) ask for clarification, (b) identify a problem, (c) explain the problem, and (d) finally suggest possible revisions. The second phase consisted of teacher–student conferences. When analysing the amount of peer feedback on both local and global issues of a peer's draft generated before and after training, it became clear that students made more comments addressing the second and third step of the 4-step strategy and also commented more on higher-order concerns.

Min's (2005) study confirms that without instruction L2 revisers frequently tend to focus on local editing issues. Min's (2005) study also supports the contention that in order to be able to communicate advice to a peer about higher-order concerns in their writing, students should be able to detect a structural or content problem, identify or diagnose it and should also know how to resolve it. Thus, training in giving higher-order feedback should in the first place be instruction in three basic mental operations inherent to most revision models (Hayes, 1996; Scardamalia & Bereiter, 1983): (a) critical reading and evaluation of text and problem detection; (b) diagnosing problems, and (c) resolving them. Students' failure to address more global problems in a peer's writing can be explained by a lack of experience with revision and lack of knowledge of revision criteria and strategies.

1.1.1. Why students fail to address global higher-order concepts

Knowledge of evaluative criteria for writing is characteristic of more expert revisers. Expert revisers are also better able to read a specific text with the target audience in mind (Sommers, 1980). They therefore not only detect more structural and content problems in a text (Hayes, Flower, Schriver, Stratman, & Carey, 1987), they also have more procedures in mind to solve global textual problems. Novice writers and revisers, on the other hand, mainly "edit" a text which means that they predominantly focus on the word and sentence level when rereading a fragment of text (Cho & MacArthur, 2010; Fitzgerald, 1992; Sommers, 1980). As a result, they fail to identify higher-order errors such as organisational problems or issues of content. Their revision process mainly involves proofreading and resembles a sort of "housecleaning task" (Baker, Gersten, & Scanlon, 2002). Furthermore, less competent revisers have fewer advanced revision strategies at their command (Faigley & Witte, 1981; Sommers, 1980). They do not necessarily split the revision process into stages and have not developed a hierarchical order of evaluative criteria.

Second (L2) or foreign language (FL) writers and revisers are, in a sense, similar to inexperienced and poor revisers. Even though they revise more frequently than in their mother tongue (Chenoweth & Hayes, 2001; Silva, 1993), the errors they identify are predominately surface-level problems (Stevenson, Schoonen, & De Glopper, 2006). This can be explained by a variety of factors. First of all, L2 writers are often preoccupied with the linguistic demands of a text and as a result may ignore levels of the text other than the linguistic (Broekkamp & Van den Bergh, 1996; Chenoweth & Hayes, 2001). This restriction to linguistic demands could also be due to an insufficient knowledge of the target language on the part of the L2 writer/reviser, poor writing skills and/or a lack of knowledge of task schemata (Stevenson et al., 2006). Next, as is the case for L1 novice revisers, FL writers may have limited knowledge of criteria for good writing and of the textual features of a specific genre and lack awareness of writing problems (Alamargot & Chanquoy, 2001; Graham, Schwartz, & MacArthur, 1993). A final explanation, which may even be more applicable to FL revisers, could be a lack of cognitive resources during revision (Chenoweth & Hayes, 2001; Stevenson et al., 2006). Inexperienced writers have not automatised all the sub-processes required for successful and efficient revision. Critical reading, detecting, diagnosing and modifying a text often results in cognitive overload in working memory (McCutchen, Kerr, & Francis, 1994).

1.1.2. Revision strategy instruction

From the expert-novice paradigm it becomes clear that in order to learn how to give more global feedback on peers' writing, students should be explicitly instructed in becoming better revisers. As a result, they should be able to detect more global problems in a peer's text, to diagnose the problems adequately and to suggest appropriate revisions. One approach is to teach students evaluative criteria and revision strategies since these have proved effective in guiding students in revision assignments (Hillocks, 1986; MacArthur, Graham, & Harris, 2004). Novice, less-skilled and FL writers and revisers could all benefit from procedural knowledge of the revision process when having to revise a text for higher-order concerns. In their recent review, Graham and Perin (2007) suggested that "explicitly and systematically teaching students strategies for planning, revising and/or editing text'' (p. 449), results in improved writing quality for adolescents in regular schools. Studies in which a revision strategy is explicitly instructed with normally-achieving students (Olson, 1990) and learning-disabled writers (Stoddard & MacArthur, 1993) have proven to be effective to improve students' revision skills.

1.1.3. Procedural facilitation

Procedural facilitation tools are scaffolding instruments that support student learning. These tools can range from cue cards, guidance sheets with prompting questions to graphic organisers. In writing, they are meant to remind writers of ''procedural steps … or higher-order strategies … to plan, monitor or revise their texts'' (Englert, Mariage, & Dunsmore, 2006, p. 211). They are particularly effective in teaching students criteria for reflecting on and evaluating writing, and thereby improving their revision skills (Englert et al., 2006; Graham & Perin, 2007). In peer assessment, peer review sheets are frequently used to make criteria more accessible to students and to prompt revision strategies (Min, 2006; Mittan, 1989). Procedural facilitators have proven effective in helping students learn to analyse texts and in considering these from the reader's perspective (Bereiter & Scardamalia, 1987); also, in making more meaningful revisions with a significant impact on text quality (Graham, 1997 with learning-disabled students; Zellermayer, Olshtain, & Cohen, 1991 with pre-academic L1 students).

1.1.4. Modelling and observational learning

One form of providing procedural facilitation that directly aims at overcoming the cognitive constraints and attention resource problems of less-skilled revisers is modelling. Modelling, which encourages the process of learning through the observing of others, is a successful instructional strategy in writing compared to more traditional methods of composition instruction which emphasise a lot of practice (Rijlaarsdam et al., 2005; Zimmerman & Kitsantas, 2002). When performing a new or cognitively complex task such as writing and/or revising, learners constantly have to grapple with learning about the task on the one hand and doing (monitoring) the task on the other. This dual agenda (Rijlaarsdam & Couzijn, 2000b) may lead to cognitive overload or a predominant focus on the production process (instead of the learning process). In the observer–learner mode (i.e., modelling) novice revisers can free more cognitive resources for the learning process itself and for the acquisition of evaluative criteria and as a result add to their metacognitive and procedural knowledge (Rijlaarsdam et al., 2005). As a result, students instructed in revision through modelling could potentially coordinate their revision process much better and be less impeded by linguistic and cognitive constraints, thus having sufficient cognitive resources to notice textual problems.

Modelling, or its desired result – observational learning – is the key factor for cognitive skill acquisition in Schunk and Zimmerman's social cognitive model of sequential skill acquisition (Schunk & Zimmerman, 1997) in which four sequential stages leading up to acquisition can be discerned: observational learning, emulation, self-control and self-regulation. The first two stages are especially informative for the design of the present study. In a first stage students observe a model, either a coping or a mastery model. This ideally results in observational learning on the part of the students. Having observed models, students imitate (emulate) the model's behaviour or ''the general form of a model's skill'' (Zimmerman & Kitsantas, 2002, p. 660). Zimmerman and Kitsantas (2002) stress that, for emulation to take place, students should ideally receive social feedback from others about their performance. In a third stage, observational and emulative learning leads to self-control by the learner, automaticity of the cognitive skill, and ultimately to self-regulation. In this model both vicarious learning, characterised by the observation of model performances and skilful behaviour, and enactive learning, learning through repeated practising until acquisition and ''skilful performance'' (Schunk, 1989), are combined.

If modelling is followed by an emulative practice session as well as social feedback, then it can be a very powerful method to acquire new, complex cognitive skills, such as writing, revision and learning the different steps in a strategy compared to more traditional methods of practising only. This became clear from Zimmerman and Kitsantas's (2002) experiment, set up to determine the effect of modelling vs. a practice-only condition without modelling on the acquisition of the complex skill of sentence combining. Moreover, the authors also wanted to test the additive value of social feedback, which was defined as feedback on students' performance provided by others (in this case, experimenters). Their hypothesis was that social feedback would add to the effect of modelling and further ensure the acquisition of a 5-step sentence-combining strategy through enhancing student motivation, which was indeed the case. Results of the pretest/posttest study also showed significant effects of modelling on sentence-combining skills.

The finding that observation offers a benefit over practising-only methods for more complicated authentic tasks such as argumentative writing and reading was also illustrated by Couzijn (1999) in an experiment with 120 Dutch students. The ''learning-by-doing groups'', as Couzijn (1999) called the two practising-only conditions, were expected to underperform because of the high cognitive load of simultaneously juggling executive processes and learning processes. This hypothesis was confirmed: significant effects for the learning-by-observation conditions on both reading and writing were found. Couzijn (1999) concluded by recommending observation for knowledge and skill acquisition followed by learning-by-doing to ''proceduralise the knowledge and making it more flexible'' (p. 134).

1.1.5. Collaboration

Finally, the effectiveness of both procedural facilitation and strategy instruction increases through social interaction and collaboration. Graham and Perin (2007) advocate ''instructional arrangements in which adolescents work together to plan, draft, revise and edit their compositions'' (p. 466), because they have a powerful effect on text quality. Baker et al. (2002) highlighted that "procedural facilitators are particularly effective when linked to dialogic interactions with others about the content and quality of the written text, as well as writing techniques and processes" (Englert et al., 2006, p. 213). Englert et al. (2006) even state that procedural facilitators are 'ineffective' without student-to-student dialogues. Interaction may also be important in observational learning. As noted above, Zimmerman and Kitsantas (2002) stress that, ideally, for emulation to take place following the observational learning phase, students should receive feedback from others about their performance. When engaged in a collaborative revision session peers can comment on each other's performance and provide feedback.

The benefits of social interaction for strategy instruction and procedural facilitation are similar to the advantages of collaboration for writing (Storch, 2005) and revision (Boscolo & Ascorti, 2004). Through discussions with peers, students are encouraged to reflect on their own learning process. Especially when composing or revising collaboratively, students can experience new and/or higher-order strategies and may emulate these and as a result achieve higher levels of competence (Baker et al., 2002; Hillocks, 1986). Through negotiation, questioning and explanation (verbalisation), students can become aware of their own strategies (Englert et al., 2006) and/or internalise the revision strategy to a greater degree. Especially since novice and inexperienced FL revisers do not detect many problems on the global textual level and fail to come up with adequate revision possibilities, they could benefit from reading and discussing texts together ("individualisation of help"; Yarrow & Topping, 2001) and also from being confronted with alternative viewpoints on a piece of writing. Thus, collaborative problem-solving not only engages students in the revision task but it may also facilitate problem detection and revision as two students may notice additional errors and different kinds of error and may produce better revision possibilities (Stoddard & MacArthur, 1993; Zammuner, 1995). Studies by Boscolo and Ascorti (2004) and Rouiller (2004) in L1 show that collaborative revision may lead to better individual revision ability (group-to-individual transfer) which is reflected in more global revisions and/or better text quality. Furthermore, the presence of a peer could reduce processing and its cognitive load as labour is divided by joint regulation (cf. De Guerrero & Villamil, 1994; Kirschner, Paas, & Kirschner, 2009; Topping & Ehly, 1998).

As the effect of observation on (subsequent) collaborative emulation in revision has rarely been investigated and as research on the combination and sequence of the different instructional components (such as collaborative or individual work) of highly effective multi-component strategy instruction packages is scarce (De La Paz, 2007), we were interested in the question whether observation or practising should be followed by collaborative or individual emulation in revision.

1.2. Research questions – hypotheses

To participate in peer revision effectively, students should be instructed to provide adequate feedback, and therefore, to revise a peer's text. This requires, first, the knowledge and skill to analyse, evaluate, and revise a text. Then it requires the skill of conveying the message to the author. In this study the focus was not so much on the tone of feedback (as is the case in Gielen, Peeters, Dochy, Onghena, & Struyven, 2010; Strijbos, Narciss, & Dünnebier, 2010) but rather on the quality of the revision, as a prerequisite for the informativeness and specificity of the feedback given. Adequate feedback here is detailed feedback which addresses global concerns in a text, uses metalanguage[1] to diagnose textual problems and suggests specific revisions. To give adequate feedback, students should be taught how to detect, discuss and revise higher-order problems in a peer's text. That is why we considered it crucial for students who give either written or oral feedback on peer-writing to first receive instruction in how to become a better reviser.

The following research questions were posed: (a) Is observation (as part of modelling) a more effective instructional strategy for students acquiring evaluative revision criteria and a revision strategy than traditional practising? (b) Is dyadic emulation more effective than individual, when exercising the procedure/strategy learnt? These two research questions led to a third one: (c) What is the effect of the combination of instruction and emulation on feedback quality. More specifically, is the sequence of observation and emulation more effective for feedback quality than the sequence practising and emulation? The following hypotheses were tested in a quasi-experimental study:

Effect of instruction. It was expected that students who acquired the revision strategy and the evaluative revision criteria through observation would outperform the students who acquired the strategy and criteria through practising (Hypothesis 1).

Effect of collaboration in emulation. It was hypothesised that after instruction, emulating the criteria in dyads would have a stronger effect on acquiring the criteria than emulating individually (Hypothesis 2).

Effect of the combination of instruction and emulation. Following on from the above two hypotheses, it was predicted that the combination of observation and dyadic emulation would be more effective for feedback quality than the combination practising and individual emulation (Hypothesis 3).

2. Method

2.1. Participants

Participants were 247 Belgian university Business freshmen (42% female) from 10 classes, randomly formed at the start of the academic year by the university administration. All students were native speakers of Dutch who took the same Business English course. Repeat students, bilingual students and students who, as a result of a one-year study abroad, had a near-native

[1] "Metalanguage" refers to the specific language to talk about problems, to define, and to diagnose them.

level of English were excluded from the sample. All students were freshmen born in the same year. Mean age of the participants was 18.5 years ($SD = 3.17$). Scores on the Oxford computer-based Quick Placement Test (QPT; Oxford University Press, 2001) showed that students were predominantly upper-intermediate EFL learners. Specifically, 36.5% of the participants had an upper-intermediate level of English, 29.1% an advanced level, and 23% a very advanced level. Of the total sample, 10.9% of the students had a lower-intermediate level. Only one student (0.4%) was an elementary EFL student.

2.2. Design – procedure

A two by two factorial design with pretest/posttest was implemented in a 6-week intervention. The factors were Instruction (Observation vs. Practising) and Emulation (Dyadic vs. Individual). Each class was randomly assigned to one of the four experimental conditions.

First, the two types of instruction aimed at instructing the same revision strategy and higher-order criteria to evaluate writing. These criteria and the different steps to take when evaluating structure and content of a peer's text were formulated as questions and revision tips on a detection and revision facilitation sheet (procedural facilitator/strategy tool; see Appendix A). In the Observation condition (O-condition), students individually observed a "mastery model" from a CD-ROM. Specifically, two expert peers modelled in dyadic interaction the application of the proposed revision strategy and criteria for revision (structure and content) to a peer's English text. The expert peers were fairly exhaustive in the discussion of the different steps of the strategy and criteria. They identified higher-order problems in the peer's text, commented and reflected on those flaws (in Dutch), suggested possible strategies to revise and corrected the flaws (in English).

In the Practising condition (P-condition) students, who worked in randomly formed dyads, acquired the revision strategy and the criteria for revision by applying them to the structure and content of the same letter discussed in the O-condition. The language of communication in the dyads was Dutch to minimise cognitive load and to stimulate more reflection (De Guerrero & Villamil, 1994; Dicamilla & Anton, 1997). These participants received a handout with the same steps of the strategy students modelled in the video and an indication of the time allocated to every single step. Instructors acted as time-keepers. After each of the six steps of the strategy, students got to see a model answer which was identical to that modelled by the dyad in the O-condition in order to ensure that potential differences between conditions cannot simply be reduced to differences in instructed content.

Time-on-task in both conditions was kept equal, that is, both conditions took 35 min.

In both conditions the instruction was followed by emulation, in which students exercised the criteria for giving feedback. There were two types of emulation, that is, students either exercised in dyads or individually. Students in the dyadic conditions were randomly grouped in dyads. This meant that dyads of the P-condition either continued exercising together (dyadic) or were split up to exercise individually, depending on the emulation type they belonged to. During emulation students in all conditions were encouraged to use the same strategy tool given in the instruction. They could also use an abridged cardboard version and access it online if they preferred that.

This design resulted in four experimental conditions; specifically, observational learning with subsequent individual (OI) and dyadic (OD) type of emulation and practising with subsequent individual (PI) and dyadic (PD) type of emulation. Emulation served a double purpose, that is, for students it was an exercising phase to emulate the criteria for giving feedback. From the researcher's point of view it provided the first testing occasion (Week 5), after the pretest (Week 1), of the four experimental conditions. One week after the intervention (i.e., instruction and emulation) students' knowledge and application of the criteria was tested individually in the posttest (Week 6).

2.3. Testing materials and dependent variables

The testing materials were (a) the pretest material, (b) the emulation material, and (c) the posttest material. These materials were letters written by peers which contained flaws at the structural and content level. The flaws were typical of the specific genre of business letter written by intermediate EFL Business students. The letters were hybrid mixtures of a letter of application and a letter of enquiry to the university's board of admissions to be accepted for a summer course on a topic related to students' field of study.

The testing materials were vetted[2] by teachers of Business English and pilot-tested for appropriate length and level of difficulty. The pretest material contained 10 flaws on the structural and content level (e.g., more than one main idea in a paragraph or an unconvincing paragraph). The emulation material and individual posttest material contained 20 flaws each. The errors and discrepancies were not indicated nor highlighted in any of the testing materials. To facilitate task-execution students worked on a computer. Cronbach's alpha of both the emulation and posttest's testing materials was satisfactory (Cronbach's $\alpha = .89$ for the emulation material and .75 for the posttest material), while for the pretest testing material Cronbach's alpha was .70, that is, acceptable.

Task instructions for the three testing materials were the following. Students were first encouraged to highlight possible flaws and inconsistencies at the structural and content level when reading the letter. Second, where possible, they needed to identify (clarify) with Word's "Comments" function (Microsoft Word 2002) what exactly the problem was in Dutch and revise in English. During emulation, since students could consult their strategy tools, the codes they used on the tool were also acceptable as comments. For the pretest and posttest testing

[2] Teachers read, checked the tests and corrected them themselves for internal inconsistencies, flaws, typos and to determine the appropriate length and level of difficulty.

materials, students were told that these comments were meant to clarify their revisions to the author. On the basis of these instructions, students received the following three scores: (a) A *detection score*, that is, a score for the number of flaws detected out of a total of 20 higher-order errors. (b) A *revision* and error-instigated corrective feedback score, that is, each student received a score for each correctly and completely revised flaw. Revision possibilities included additions (adding information to paragraphs or adding paragraphs), deletions (deleting parts of text), and reorderings (of information in the letter) (cf. Faigley & Witte, 1981). (c) A *comments* or diagnosis score (at pretest and posttest materials), that is, a holistic score for the correctness, exhaustiveness and explicitness of student comments ranging from 1 to 5. In this score we included the way in which students identified and discussed the error, suggested revisions to the author or commented on the revisions they made. The scoring procedure was the following. On the basis of a first screening of a large part of the pretest and posttest comments, the first author, in agreement with experienced teachers of the specific genre, proposed a 5-point scale and decided on examples for the different scores depending on comments' level of correctness and exhaustiveness. On the basis of these examples the remaining test comments were scored by the researcher or first author and 10% additionally by an experienced rater.

Quality of peer feedback in the present study was, therefore, indicated by three variables: (a) detection, (b) revision, that is, the quality of corrective feedback of the detected errors, and (c) comments, namely the quality of the comments that students inserted. In detections, revisions (transformations and changes), and comments the way in which students applied the different evaluative criteria and the revision strategy were investigated. A second experienced rater, blind to the design of the study, scored 10% of all the tests ($n = 60$). Correlation between the two sets of independent scores (first author's and rater's) was high for detection ($r = .84$, $p < .05$), revision ($r = .80$, $p < .05$), and comments ($r = .85$, $p < .05$). Prior to the analyses, differing scores were discussed until agreement was reached.

2.4. Data analyses

First of all, it should be noted that the number of students varied from 235 to 247. All students were present during the instruction and emulation sessions but for the pretest and posttest materials there were a few missing values.

Because of the specific nature of the data collected in this study, multivariate multilevel analyses with two fixed factors, Instruction and Emulation, were used. Each factor had two levels; more specifically, Instruction (i.e., Factor 1) was represented by Observation or Practising level, and Emulation (i.e., Factor 2) by the Individual or the Dyadic level.

The data were hierarchically structured in the following way. The measures of the pretest, posttest, and emulation were included in the model (e.g., detection, revision, comments). These measures were nested within individuals. Hence, the model applied was a multivariate multilevel fixed occasion model (Goldstein, 2003) with two levels, namely measures at the first level, and individuals at the second level (Hox, 2002). For each measure in each experimental condition the variance as well as the mean was estimated. The MLwiN 2.10 software (Rasbash, Charlton, Browne, Healy, & Cameron, 2009) was used to conduct the analyses.

3. Results

3.1. Preliminary analyses

As classes were randomly assigned to the four different conditions, systematic differences in language proficiency (on QPT) between them were unlikely and were not found, $F(3, 235) = 1.24$, $p = .271$, partial $\eta^2 = .05$. No initial differences in proficiency in English on the QPT existed between the four different conditions either, $F(3, 230) = 1.03$, $p = .382$, partial $\eta^2 = .01$. The distribution of male and female students was not significantly different over conditions, $\chi^2(3, N = 247) = 1.20$, $p = .754$. The scores on the pretest material showed that the majority of the participants were novice revisers in English (Table 1).

3.2. The multivariate multilevel model

The multivariate multilevel fixed occasion model tested is explained below. As clarified in Section 2.3 there were three testing materials (pretest, emulation, and posttest) each with different measures (detection, revision, comments). Hence, in the multivariate model all students took eight test

Table 1
Mean (and standard deviations) for conditions for pretest measures.

	Conditions				
	Observation		Practice		
	Dyad (OD)	Individual (OI)	Dyad (PD)	Individual (PI)	
Measures	($n = 88$)	($n = 44$)	($n = 72$)	($n = 43$)	Total mean
QPT[a]	72.33 (10.01)	72.65 (10.12)	69.83 (9.72)	70.27 (11.03)	71.23 (10.15)
Detection[b]	0.80 (0.13)	1.24 (0.19)	0.59 (0.13)	0.99 (0.18)	0.91 (0.16)
Revision[b]	0.37 (0.94)	0.47 (1.18)	0.82 (0.37)	0.45 (1.25)	0.35 (0.95)
Comments	0.31 (0.06)	0.40 (0.08)	0.09 (0.07)	0.39 (0.08)	0.32 (0.07)

For the total number of dyads in OD and PD, divide n by 2.
[a] QPT is scored out of 100.
[b] Pretest detection and revision scores are rescaled/recalibrated scores out of 20 items.

Table 2
Estimated covariance matrix (diagonal and below) between the eight measures (standard errors in parenthesis) and correlation matrix (above diagonal).

	Measures							
	P1	P2	P3	E1	E2	T1	T2	T3
Pretest detection (P1)	1.38 (0.13)	0.61	0.63	−0.70	−0.08	0.09	0.04	0.02
Pretest revision (P2)	0.67 (0.09)	0.88 (0.08)	0.45	−0.05	−0.03	0.07	0.05	0.14
Pretest comments (P3)	0.50 (0.06)	0.29 (0.05)	0.45 (0.04)	−0.00	−0.11	0.05	0.04	0.01
Emulation detection (E1)	−2.11 (0.21)	−0.11 (0.16)	−0.08 (0.12)	6.57 (0.63)	0.81	0.37	0.37	0.30
Emulation revision (E2)	−0.66 (0.59)	−0.19 (0.47)	−0.55 (0.34)	15.32 (1.65)	54.69 (5.25)	0.29	0.34	0.29
Posttest detection (T1)	0.29 (0.22)	0.19 (0.18)	0.09 (0.13)	2.63 (0.52)	5.94 (1.45)	7.74 (0.74)	0.79	0.36
Posttest revision (T2)	0.27 (0.43)	0.25 (0.34)	0.14 (0.24)	5.03 (0.99)	13.28 (2.83)	11.70 (1.28)	28.51 (2.74)	0.28
Posttest comments (T3)	0.03 (0.09)	0.16 (0.08)	0.00 (0.05)	0.93 (0.22)	2.57 (0.62)	1.20 (0.24)	1.79 (0.45)	1.42 (0.14)

measures or tests; specifically, a pretest with three measures, detection, revision and comments (represented by P1, P2, P3), emulation with two measures, detection and revision (represented by E1, E2), and three posttest measures, detection, revision, and comments (represented by T1, T2, T3). If Y_{ij} is the score of student j ($j = 1, 2, \ldots, n$) on the i ($i = 1, 2, \ldots, 8$) test, and the conditions are defined by means of dummy variables, say $C1_{ij}$, $C2_{ij}$, $C3_{ij}$ and $C4_{ij}$, the model to be analysed can be written as:

$$\begin{aligned} Y_{ij} = {} & P1_{ij}\left(\beta_1 \times C1_{ij} + \beta_2 \times C2_{ij} + \beta_3 \times C3_{ij} + \beta_4 \times C4_{ij} + e_{1j}\right) \\ & + P2_{ij}\left(\beta_5 \times C1_{ij} + \beta_6 \times C2_{ij} + \beta_7 \times C3_{ij} + \beta_8 \times C4_{ij} \right. \\ & \left. + e_{2j}\right) + P3_{ij}(\ \ldots\ ) + E1_{ij}(\ \ldots\ ) + E2_{ij}(\ \ldots\ ) \\ & + T1_{ij}(\ \ldots\ ) + T2_{ij}(\ \ldots\ ) + T3_{ij}\left(\beta_{29} \times C1_{ij} + \beta_{30} \right. \\ & \left. \times C2_{ij} + \beta_{31} \times C3_{ij} + \beta_{32} \times C4_{ij} + e_{8j}\right) \end{aligned} \tag{1}$$

In Equation (1) $P1_{ij}$, $P2_{ij}$,… $T3_{ij}$ are dummy variables, denoting the eight tests or test measures. Hence, the regression weights (β_1, β_2,… β_{32}) are the estimated means of each test in each condition (see Table 3). The differences between conditions (for each test measure) can be estimated by means of a contrast procedure (Goldstein, 2003; Snijders & Bosker, 1999).

The residual scores (e_{1j}, e_{2j},… e_{8j}) represent the deviation from the mean score. All residuals are assumed to be normally distributed with a variance of S^2_{e1}, S^2_{e2},… S^2_{e8}, and an expected value of zero. As all eight tests or test measures are taken by the same individuals, the eight residuals are allowed to co-vary (see Table 2).[3]

3.2.1. Pretest measures

Analyses of the scores in the pretest showed a floor effect. Specifically, 80% of all students scored zero for the detection and revision of structure and content flaws which confirms that they were predominantly novice revisers of English. This zero score does not necessarily mean that students did not detect, revise, or comment at all but that they did not notice, identify, or revise global issues such as structure and content. No initial differences between the different experimental conditions could be observed for detection, $\chi^2(6, N = 235) = 8.11$, $p = .23$, revision, $\chi^2(6, N = 235) = 4.13$, $p = .67$, or comments, $\chi^2(6, N = 235) = 11.96$, $p = .063$.

3.2.2. Emulation and posttest measures

The percentage of variance the multilevel fixed occasion model explained for the measures was 12% for emulation detection, 6% for emulation revision, 11% for posttest detection, 9% for posttest revision and 9% for posttest comments. According to Cohen (1977) these are medium effect sizes. Table 3 illustrates the descriptives for the emulation and posttest measures.

3.2.3. Effects of instruction type and emulation type on emulation scores

A main effect of the instruction type was observed for detection only, $\chi^2(1, N = 247) = 8.52$, $p = .003$, with Observation outperforming Practising. No main effects for detection, $\chi^2(1, N = 247) = 0.45$, $p = .502$, or revision, $\chi^2(1, N = 247) = 0.01$, $p = .920$ were found for emulation type. However, as illustrated in Fig. 1 for revision only, we did observe a statistically significant interaction effect for both detection, $\chi^2(1, N = 247) = 9.67$, $p = .001$, and revision,

Table 3
Means (and standard errors) of the dependent variables in the emulation and posttest.

Measures	Conditions			
	Observation		Practice	
	Dyad (OD)	Individual (OI)	Dyad (PD)	Individual (PI)
Emulation				
Detection	12.01 (0.29)	11.13 (0.41)	9.83 (0.33)	11.20 (0.40)
Revision	19.34 (0.79)	15.50 (1.15)	15.47 (0.88)	19.12 (1.13)
Posttest				
Detection	8.07 (0.31)	7.99 (0.44)	6.08 (0.35)	8.18 (0.44)
Revision	9.19 (0.58)	9.86 (0.82)	6.22 (0.65)	9.93 (0.83)
Comments	2.87 (0.27)	2.75 (0.38)	2.74 (0.30)	3.77 (0.40)

[3] For nested models where one model contains another MLwiN provides the −2 log likelihood as an indicator of the fit of the model instead of BIC and AIC measures. For normally distributed data the −2 log likelihood statistic can be used when two or more models are compared. However, as we do not compare different models, the −2 log likelihood or deviance statistic is not provided.

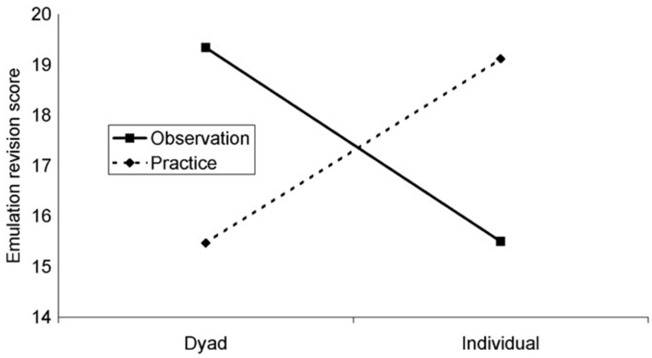

Fig. 1. Interaction of instruction with emulation for revision in the emulation.

$\chi^2(1, N = 247) = 14.19$, $p < .001$. Observation was more effective with Dyadic Emulation, whereas Practice was more effective with Individual Emulation.

3.2.4. *Effects of instruction type and emulation type on posttest scores*

The results showed a statistically significant main effect of both the instruction type and the emulation type for detection, $\chi^2(1, N = 238) = 5.35$, $p = .020$, and $\chi^2(1, N = 238) = 6.78$, $p < .001$, respectively, and for revision, $\chi^2(1, N = 238) = 4.01$, $p = .045$, and $\chi^2(1, N = 238) = 9.09$, $p < .001$, respectively. Students in the observational learning condition were better at detecting, identifying higher-order problems and applying the evaluative criteria to a peer's text than students in the practising learning condition (with a peer). Also, individuals seemed to do better than dyads. However, these two main effects have to be put into perspective because of a statistically significant interaction of instruction type with emulation type. This interaction could be observed for detection, $\chi^2(1, N = 238) = 7.85$, $p < .001$ (see Fig. 2), and revision, $\chi^2(1, N = 238) = 4.41$, $p = .035$ (see Fig. 3).

These findings again reveal that the combination of practice (as instruction) and dyad (as emulation) is not an effective instructional strategy. As far as detection is concerned, the two most successful conditions were PI and OD. The two most powerful conditions for revision were PI and OI. For both detection and revision emulation in dyads seems to be productive if preceded by observational learning only. If students acquire the criteria for giving feedback by practising, emulation in dyads seems to be least effective. As becomes apparent from Table 3, the PD condition had least of all conditions internalised the criteria for giving feedback on structure and content of a peer's text.

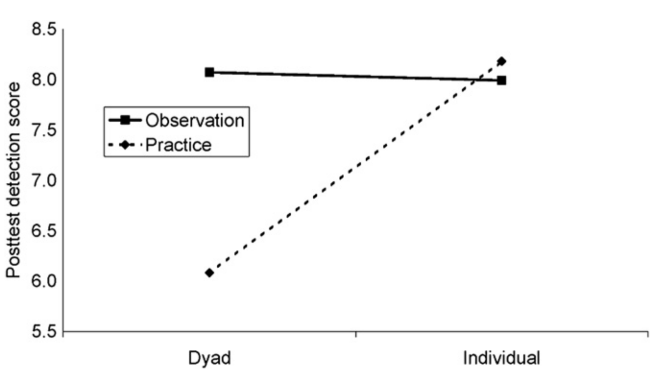

Fig. 2. Interaction of instruction with emulation for detection in the posttest.

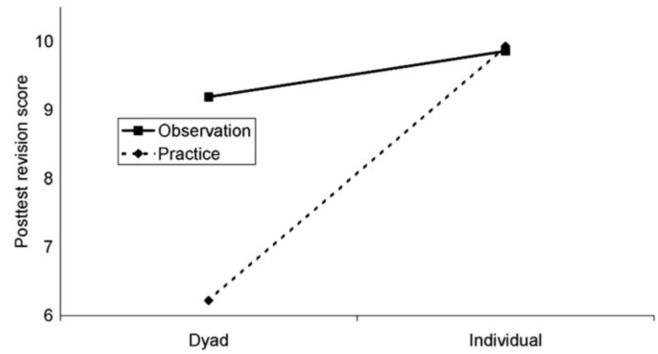

Fig. 3. Interaction of instruction with emulation for revision in the posttest.

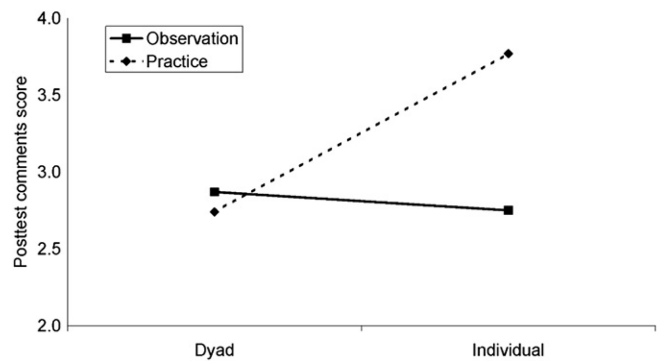

Fig. 4. Interaction of instruction with emulation for comments in the posttest.

The above results were further corroborated in the case of comments in the posttest where a statistically significant main effect of emulation type for comments was found, $\chi^2(1, N = 236) = 6.63$, $p < .001$ (Fig. 4).

This effect was moderated by the Instruction factor as the significant interaction of instruction with emulation indicated, that is, OD and PI turned out to be the stronger conditions, $\chi^2(1, N = 236) = 7.68$, $p < .001$.

4. Discussion

Results on the pretest indicated that without instruction novice revisers were not able to do more than superficial grooming when asked to revise a peer's draft. Analyses of the detections and revisions students made revealed that they rarely considered the text holistically and mainly focused on the word and sentence level. In that respect, without instruction, corrective feedback and comments made by these EFL novice revisers may be helpful for their peers to edit their writing but do not necessarily help them to improve the structure or content of their drafts. To enable EFL students to also provide their peers with meaningful feedback on global issues in their writing, they should first be taught how to detect global flaws and secondly how to remedy these flaws. In other words, novice revisers should be instructed in how to become better revisers in order to give correct, explicit and specific feedback which surpasses the level of so-called rubber-stamp advice (Leki, 1990).

Observational learning was hypothesised to create an effective learning environment for novice revisers to become better revisers. This would be achieved by splitting up learning and task-execution, in order to reduce the load of cognitive resources on the part of the learner. Observation of two expert peer revisers was assumed to enable the novice revisers in this study to first learn a revision strategy and evaluative criteria for writing after which an emulative application phase with a peer would further enhance the internalisation of these criteria. As a result, participants would be able to give more correct and explicit corrective feedback when evaluating a peer's text. Learning about the criteria by applying them with a peer to a text (practising condition) on the other hand, was considered to be potentially less effective due to the dual agenda of the learner–reviser who at the same time had to learn the revision strategy (=learning) and do the complicated task of reading a draft, reflecting and commenting on it and identifying flaws. The simultaneous juggling of the roles of reader–reviser and learner could lead to less focus on the learning process and ultimately also to too many cognitive constraints and possibly cognitive overload. However, collaboration was supposed to overcome possible cognitive constraints and to facilitate the acquisition of the strategy. The results of the present study urge us to adjust our assumptions.

As regards the first research question whether observational learning is a more effective instructional strategy for acquiring the evaluative criteria and the strategy than traditional practising, the answer is not so straight-forward. The effect of instruction depends on the setting of the subsequent type of emulation. Observation proves to be a powerful instructional strategy on the condition that consequent emulation is a collaborative undertaking. However, a more traditional practice-only instruction followed by individual emulation appears to be as productive or even more (posttest revision score). We can thus conclude that Hypothesis 1 was falsified.

Equally, the answer to the second research question whether emulating the evaluative criteria in a dyad after instruction will make students internalise the criteria to a more significant degree, depends on the preceding instruction. Apparently, if students acquire the evaluative criteria in a more traditional way, that is, by applying them, internalisation is less likely to happen if students subsequently emulate them with a peer rather than individually. Indeed, the presence of a peer in the PD condition did not compensate for the lack of observational learning. On the contrary, students without the aid of a peer acquired the revision criteria better and the PD condition turned out to be quite ineffective as far as skill and strategy acquisition are concerned. If, however, collaborative emulation is preceded by observational learning, students can benefit significantly from peer interaction and elaborated dialogue. Hence, Hypothesis 2 was only partly confirmed.

Following from the answers to the first two research questions it becomes clear that the answer to the third question as to which combination of instruction and emulation is most effective for the acquisition and application of revision criteria is not that straight-forward. Contrary to our expectations (cf. Hypothesis 3), OD does not seem to be the most effective condition, even though the results show that observation is an effective instruction method. For the practising condition, however, whether or not to have students emulate in dyads or individually post instruction turns out to be a crucial decision, as the posttest results revealed. Practising followed by dyadic emulation seems to be the least successful combination, whereas practising followed by individual emulation appears to be as effective as dyadic observation.

A few questions need to be addressed. First of all: why do the observation conditions not outperform all practice conditions? Experimental studies by Braaksma, Rijlaarsdam, and Van den Bergh (2002) showed that individual differences play a significant role in the effectiveness of observational learning. Possibly, an interaction effect could be observed between a students' level of writing proficiency and the treatment. Further investigations need to be conducted to look at the interplay of individual student characteristics and the instructional strategies investigated in the present study.

Second, the fact that collaboration in emulation is not always an added value has been pointed out by researchers in the collaborative learning field. They stress that collaborative problem-solving is not tantamount to cognitive growth (Barron, 2003). To ensure the triggering of learning mechanisms and the acquisition of cognitive skills, a lot of attention needs to be paid to implementing the right scaffolds (Dillenbourg, 1999). It appears that observational learning is such a scaffold as opposed to practising. Barron (2003) stresses the crucial role of "the achievement of joint attention (…) for problem-solving and learning" (p. 310). Modelling may be a more effective instructional strategy to guarantee joint attention as opposed to practising. It may prepare dyads better for yet another dual agenda of simultaneous problem-solving and interacting, which Barron (2003) refers to as "the dual-space demands of collaboration" (p. 353). Possibly, the cognitive load for students in the dyadic practice condition was too high and instead of alleviating the cognitive burden of the task-execution phase, the peer added to it (cf. Dillenbourg, 1999). Individual students in the practice condition possibly had more available attentional and cognitive resources to learn about the strategy and reflect on the criteria. Further research needs to be conducted to empirically verify these hypotheses. Future experiments could include a cognitive load test to verify if the difference between the conditions can indeed be explained in terms of attentional resources available to the participants (Kirschner et al., 2009).

Finally, it needs to be stressed that there is no indication whatsoever that students in dyads in the PD condition did not take the task seriously or that more off-task talk was observed in the dyads in this condition as opposed to the other experimental groups. Nevertheless, due to the complicated nature of collaborative learning, further investigations of what exactly went on in the dyads in the PD condition as opposed to the dyads in the OD condition may further clarify the findings. To understand more fully why some dyads revise better than others the quantitative results presented in this study should be complemented with a qualitative analysis of student interaction and group dynamics in the dyads. In this way the impact of both group configuration and group interaction on revision products could possibly explain the results in more detail (cf. Dillenbourg, 1999). The different data sources collected and a combination of both qualitative and quantitative measures (triangulation) will enable us to verify, explore and fine-tune the findings of this study.

To conclude, an important caveat needs to be noted regarding the present study. Whereas most studies on peer response offer quite substantive and prolonged training, this study was set up not with the aim of looking at long-term effects but at investigating the most effective way of instructing students to become better feedback providers. Further research needs to explore whether these effects are retained in time.

Nevertheless, one of the strengths of this study is that it showed the importance of studying the effect of combinations of subsequent instructional methods and emulation activities (i.e., learning and exercising activities), so-called more ‘hybrid’ methods (cf. Baker et al., 2002). In most experimental studies on strategy instruction, a complete programme is compared to another programme with differences in various components. For proper theory-building, however, we should break the programmes down into components and not only study the effect of the components but also of their combination and sequence. The findings of this study contribute to our understanding of the impact of observational learning, practising and dyadic collaboration for revision and serve as a basis for further research studies and theory-building. A complete theory of writing instruction should not only take into account (a) the instructional components, (b) their combination and sequence, and (c) relevant design parameters such as individual or collaborative work but also (d) the interaction between instructional components and learner variables on the one hand (cf. Braaksma et al., 2002; Kieft, Rijlaarsdam, & Van den Bergh, 2008) and between learner variables, task features, group composition and interaction on the other (cf. Dillenbourg, 1999).

Implications for teaching practice seem to be that for collaborative revision to be effective as a method for learning criteria for writing, it needs to be preceded by observation. The data in this study suggest that having two students revise together with the aid of a strategy tool (procedural facilitator) seems to be not so effective. Indeed, the procedural facilitator if not modelled beforehand, seems to be a better instructional strategy for individual revision. Future research should verify if these findings are also upheld for other writing tasks.

Appendix A. A step on the strategy tool

I. FOCUS on CONTENT of the LETTER as a whole

> Step 1: Is the letter CONVINCING?
>
> “If you were responsible for accepting students for this particular summer school, would you accept the writer of this letter for this particular course?” *Yes/No*

→ IF NOT

> “Is it because the letter is not complete?” *Yes/No*
>
> “Would you accept the writer on the basis of the amount of information mentioned in the letter? Does the writer include sufficient paragraphs to reach this goal?” *Yes/No*
>
> 1) Is the aim of writing clearly mentioned? Is the purpose of writing immediately clear so that the reader can classify the letter on a specific pile? *Yes/No**
>
> 2) Does the writer introduce him/herself? *Yes/No**
>
> 3) Does the writer motivate why he/she would like to participate in the course? *Yes/No**
>
> 4) Does the writer prove that he/she is a suitable applicant? *Yes/No**
>
> 5) Does the writer ask practical questions and additional info? *Yes/No**

*IF the answer to one or more of these questions is “No” ⇒ *Revision tips*: Add the missing information.

References

Alamargot, D., & Chanquoy, L. (2001). *Through the models of writing*. Dordrecht, the Netherlands: Kluwer.

Baker, S., Gersten, R., & Scanlon, D. (2002). Procedural facilitators and cognitive strategies: tool for unravelling the mysteries of comprehension and the writing process, and for providing meaningful access to the general curriculum. *Learning Disabilities: Research and Practice, 17*, 65–77.

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*, 307–359.

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

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

Boscolo, P., & Ascorti, K. (2004). Effects of collaborative revision on children’s ability to write understandable narrative texts. In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes*. In

Rijlaarsdam, G. (Ed.), *Studies in writing, Vol. 13* (pp. 157–170). Dordrecht, the Netherlands: Kluwer.

Braaksma, M., Rijlaarsdam, G., & Van den Bergh, H. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology, 94*, 405–415.

Broekkamp, H., & Van den Bergh, H. (1996). Attention strategies in revising a foreign language text. In G. Rijlaarsdam, H. Van den Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 170–181). Amsterdam: Amsterdam University Press.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: generating text in L1 and L2. *Written Communication, 18*, 80–98.

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

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* [Revised edition]. New York: Academic.

Couzijn, M. (1999). Learning to write by observation of writing and reading processes: effects on learning and transfer. *Learning and Instruction, 9*, 109–142.

De Guerrero, M. C. M., & Villamil, O. (1994). Social-cognitive dimensions of interaction in L2 peer revision. *The Modern Language Journal, 78*, 484–496.

De La Paz, S. (2007). Managing cognitive demands for writing: comparing the effects of instructional components in strategy instruction. *Reading and Writing Quarterly, 23*, 249–266.

Dicamilla, F. J., & Anton, M. (1997). Repetition in the collaborative discourse of L2 learners: a Vygotskyan perspective. *Canadian Modern Language Review, 53*, 609–633.

Dillenbourg, P. (1999). What do you mean by 'collaborative leaning'? In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp. 1–19). Oxford: Elsevier.

Englert, C., Mariage, T. V., & Dunsmore, K. (2006). Tenets of sociocultural theory in writing instruction research. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 208–221). New York: Guilford.

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

Fitzgerald, J. (1992). *Towards knowledge in writing: Illustration from revision studies*. New York: Springer.

Flynn, E. (1982, November). *Effects of peer critiquing and model analysis on the quality of biology student laboratory reports.* Paper presented at the annual meeting of the National Council of Teachers of English, Washington, DC, USA (ERIC Document Reproduction Service No. ED 234 403).

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.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Arnold.

Graham, S. (1997). Executive control in the revising of students with learning and writing difficulties. *Journal of Educational Psychology, 89*, 223–234.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*, 445–476.

Graham, S., Schwartz, S., & MacArthur, C. (1993). Learning disabled and normally achieving students' knowledge of the writing and the composing process, attitude to writing and self-efficacy. *Journal of Learning Disabilities, 26*, 237–249.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy, & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 1–27). Mahwah, NJ: Erlbaum.

Hayes, J. R., Flower, L. S., Schriver, K. A., Stratman, J., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Reading, writing, and language processing. Advances in psycholinguistics, Vol. 2* (pp. 176–240). Cambridge, UK: Cambridge University Press.

Hillocks, G., Jr. (1986). *Research on written composition: New directions for teaching*. Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.

Hox, J. (2002). *Multilevel analysis: techniques and applications*. Mahwah, NJ: Erlbaum.

Kieft, M., Rijlaarsdam, G., & Van den Bergh, H. (2008). An aptitude–treatment interaction approach to writing-to-learn. *Learning and Instruction, 18*, 379–390.

Kirschner, F., Paas, F., & Kirschner, P. A. (2009). Individual and group-based learning from complex cognitive tasks: effects on retention and transfer. *Computers in Human Behavior, 25*, 306–314.

Leki, I. (1990). Potential problems with peer responding in ESL writing classes. *CATESOL Journal, 3*, 5–17.

MacArthur, C. A., Graham, S., & Harris, K. R. (2004). Revision with struggling writers. In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes*. In Rijlaarsdam, G. (Ed.), *Studies in writing, Vol. 13* (pp. 125–137). Dordrecht, the Netherlands: Kluwer.

McCutchen, D., Kerr, S., & Francis, M. (1994, April). *Editing and revising: Effects of knowledge of topic and error location.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, USA.

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

Mendonca, C. O., & Johnson, K. E. (1994). Peer review negotiations: revision activities in ESL writing instruction. *TESOL Quarterly, 28*, 745–769.

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

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

Mittan, R. (1989). The peer review process: harnessing students' communicative power. In D. M. Johnson, & D. H. Roen (Eds.), *Richness in writing: Empowering ESL students* (pp. 207–219). New York: Longman.

Olson, V. L. B. (1990). The revising process of sixth-grade writers with and without peer feedback. *Journal of Educational Research, 84*, 22–29.

Oxford University Press. (2001). *Oxford computer-based Quick Placement Test (QPT).* Oxford, UK: Author.

Rasbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2009). *MLwiN (Version 2.10).* [Computer software]. Bristol, UK: Centre for Multilevel Modelling.

Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Kieft, M., & Broekkamp, H. (2005). Psychology and the teaching of writing in 60 minutes. In P. Tomlinson, J. Dockrell, & P. Winne (Eds.), *Pedagogy – Teaching for learning. BJEP monograph, Series II, No. 3* (pp. 127–153). Leicester, UK: The British Psychological Society.

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

Rijlaarsdam, G., & Couzijn, M. (2000b). Writing and learning-to-write. A double challenge. In R. Simons, J. Van der Linden, & T. Duffy (Eds.), *New learning* (pp. 157–190). Dordrecht, the Netherlands: Kluwer.

Rouiller, Y. (2004). Collaborative revision and metacognitive reflection in a situation of narrative text production. In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes.* In Rijlaarsdam, G. (Ed.), *Studies in writing, Vol. 13* (pp. 171–187). Dordrecht, the Netherlands: Kluwer.

Scardamalia, M., & Bereiter, C. (1983). The development of evaluative, diagnostic and remedial capabilities in children's composing. In M. Martlew (Ed.), *The psychology of written language. Developmental and educational perspectives* (pp. 67–95). New York: Wiley & Sons.

Schunk, D. H. (1989). Social cognitive theory and self-regulated learning. In D. H. Schunk, & B. J. Zimmerman (Eds.), *Self-regulated learning and academic achievement: Theory, research and practice* (pp. 83–110). London: Erlbaum.

Schunk, D. H., & Zimmerman, B. J. (1997). Social origins of self-regulatory competence. *Educational Psychologist, 32*, 195–208.

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: the ESL research and its implications. *TESOL Quarterly, 27*, 657–677.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modelling*. London: Sage.

Sommers, N. (1980). Revision strategies of student writers and experienced writers. *College Composition and Communication, 31*, 378–387.

Stanley, J. (1992). Coaching student writers to be effective peer evaluators. *Journal of Second Language Writing, 1*, 217–233.

Stevenson, M., Schoonen, R., & De Glopper, K. (2006). Revising in 2 languages: a multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing, 15*, 201–233.

Stoddard, B., & MacArthur, C. A. (1993). A peer editor strategy: guiding learning disabled students in response and revision. *Research in the Teaching of English, 27*, 76–103.

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

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

Tang, G., & Thitecott, J. (1999). Peer response in ESL writing. *TESL-Canada-Journal, 16*(2), 20–38.

Topping, K., & Ehly, S. (1998). Introduction to peer-assisted learning. In K. Topping, & S. Ehly (Eds.), *Peer-assisted learning* (pp. 1–23). Mahwah, NJ: Erlbaum.

Yarrow, F., & Topping, K. (2001). Collaborative writing: the effects of metacognitive prompting and structured peer interaction. *British Journal of Educational Psychology, 71*, 261–282.

Zammuner, V. L. (1995). Individual and cooperative computer-writing and revising: who gets the best results? *Learning and Instruction, 5*, 101–124.

Zellermayer, M., Olshtain, E., & Cohen, J. (1991). The development of elaborative skills: teaching basic writing students to make the commitment to audience and topic. *Linguistics and Education, 3*, 359–383.

Zimmerman, B., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94*, 660–668.

Zhu, W. (1995). Effects of training for peer response on students' comments and interaction. *Written Communication, 12*, 492–528.