The mediating effect of instruction on pair composition in L2 revision and writing

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

DOI
10.1007/s11251-014-9318-5

Publication date
2014

Document Version
Final published version

Published in
Instructional Science

Instr Sci (2014) 42:905–927
DOI 10.1007/s11251-014-9318-5

The mediating effect of instruction on pair composition in L2 revision and writing

Elke Van Steendam · G. C. W. Rijlaarsdam · H. H. Van den Bergh · L. Sercu

Received: 8 July 2013 / Accepted: 5 April 2014 / Published online: 4 May 2014

Abstract The present study tests the effect of ability pairing in two instructional methods in L2 collaborative revision. Two continuous indices determine a pair: individual proficiency level, distance in proficiency between pair members (heterogeneity), and the interaction between both indices. Instructional methods tested are modelling and practising. Results show that the effect of pair composition depends on instructional strategies. In the Practising condition less proficient learners profit most from a heterogeneous ability pair, whereas more proficient learners are best paired homogeneously. In the Modelling

E. Van Steendam
Faculty of Applied Economic Sciences, University of Antwerp, Antwerp, Belgium

E. Van Steendam
Faculty of Arts, KU Leuven – University of Leuven, Leuven, Belgium

Present Address:
E. Van Steendam (✉)
Faculty of Arts, KU Leuven – University of Leuven, Campus Brussels, Warmoesberg 26, 1000 Brussels, Belgium
e-mail: elke.vansteendam@telenet.be; elke.vansteendam@arts.kuleuven.be

G. C. W. Rijlaarsdam
Research Institute on Child Development and Education, University of Amsterdam, Nieuwe Prinsengracht 130, 1018 VZ Amsterdam, The Netherlands
e-mail: g.c.w.rijlaarsdam@uva.nl

G. C. W. Rijlaarsdam
Faculty of Arts, University of Antwerp, Rodestraat 14, 2000 Antwerpen, Belgium

H. H. Van den Bergh
Faculty of Humanities, Department of Languages, Literature and Communication, University of Utrecht, Trans 10, 3512JK Utrecht, The Netherlands
e-mail: h.vandenbergh@uu.nl

L. Sercu
Faculty of Arts, Department of Language and Education, KU Leuven – University of Leuven, Blijde Inkomststraat 23, 3000 Leuven, Belgium

condition no effect of pair composition factors was observed. This result illustrates that Modelling is a powerful instructional method for complex learning tasks like collaborative revision in L2 as it overrides some of the grouping effects which can be found in more traditional learning conditions.

Keywords Collaborative revision · L2-writing · Ability grouping · Observational learning · Modelling · Group composition

Conditions for successful collaboration: group composition and instruction

One expects a lot of peer collaboration for both learning-to-write and learning-to-revise in a first language (=L1) (MacArthur et al. 1991; Stoddard and MacArthur 1993) and a second or foreign language (=L2) (Villamil and De Guerrero 1996, 1998). To have a positive effect, however, peer collaboration should be structured by training or instruction (cf. Min 2005; Van Steendam et al. 2010). Another decisive factor that contributes to the effect of collaboration is group formation as studies on small group learning in mathematics and science education demonstrate (Fuchs et al. 1998; Webb et al. 2002). However, few studies in writing and revision instruction explore the effect of ability grouping (Sutherland and Topping 1999). Therefore the present study investigates the main effects and the interaction of two pair composition factors in two instructional methods for undergraduates learning-to-revise and learning-to-write in English as a foreign language (EFL). It does so by defining pair composition with the continuous indices 'individual performance level' of a pair member and 'distance in performance' between the pair members, and the interaction of these two indices. The instructional settings studied are learning-by-practising and learning-by-observation.

Ability group composition

A vast body of research exists on ability composition of small groups in collaborative learning. The majority of these studies are situated in science and mathematics education and in computer-supported collaborative learning (Hooper and Hannafin 1988; Webb et al. 1998; Webb et al. 2002). They test the impact of group configuration on a number of intermediate and dependent variables such as interaction, group functioning and performance (Hooper and Hannafin 1991; Webb et al. 2002). Participants are predominantly 6th- to 8th-grade students grouped in triads (Webb et al. 1998) or quartets (Hooper and Hannafin 1988).

These studies present conflicting views about which group composition is more effective: a homogeneous (same-ability) or heterogeneous (mixed-ability) group. The majority, however, recommend heterogeneous grouping especially for the weaker students because they learn more in heterogeneous than in homogeneous groups (Hooper and Hannafin 1988; Webb 1980; Wiedmann et al. 2012). Theoretical explanations are provided by Piaget's concept of sociocognitive conflict (De Lisi and Golbeck 1999; Piaget 1932) and Vygotsky's Zone of Proximal Development (ZPD) (Hogan and Tudge 1999; Vygotsky 1978).

Piaget's concept of (socio)cognitive conflict posits that if students in a small group have conflicting opinions, they may question their own and each other's perspectives and as a result reorganize and add to existing knowledge. This would be highly beneficial for a successful group task (De Lisi and Golbeck 1999; Webb and Palincsar 1996) and ultimately for cognitive growth (Daiute and Dalton 1993).

For sociocognitive conflict to result in cognitive change, peers need to perceive conflicting viewpoints, which is more likely to occur in heterogeneous ability groups (Hogan and Tudge 1999; Mugny and Doise 1978). However, if the cognitive distance between the peers is too large, conflict may not occur and an alternative point of view may simply be accepted instead of argued or negotiated (Webb and Palincsar 1996).

Also in Vygotsky's ZPD, distance between group members plays an important role. Vygotsky's ZPD originally focused on parent–child interaction. It is the distance between learners' actual and the potential developmental level they can reach with the guidance (Wood et al. 1976) of a more competent peer (Hogan and Tudge 1999; Vygotsky 1978).

In mixed-ability groups, the relatively weaker student can learn from observing more competent peers' advanced processes and their explanations. Especially the mechanism of explaining, which may occur more often in heterogeneous groups (Webb et al. 2002), would result in learning for both the recipient of explanations and the student explaining, provided that the explanation is elaborate and "procedurally clear" (Fuchs et al. 1998, p. 228; Webb and Palincsar 1996).

Vygotsky's ZPD "has generally been taken to imply that neither the task difficulty nor the guidance given (…) should be too far in advance of [learners'] (…) current level of ability" (Hogan and Tudge 1999, p. 51). However, even though distance between group members in heterogeneous grouping is crucial both in Vygotsky's ZPD and in Piaget's sociocognitive conflict, it is rarely included as a separate variable to operationalize group composition. Groups are categorized as either same- or mixed-ability groups on the basis of a single, dichotomous pretest score of the group members: heterogeneous groups consist of both high- and low-achievers and homogeneous groups of same-ability students. In the few instances where a wider ability range within heterogeneous groups is explored, that is, where next to high- and low-achievers also medium-ability students are included (cf. Webb et al. 1998), defining the ability strata is also done on the basis of a single, dichotomous pretest score.

For the low-achiever to benefit from the high-achiever's scaffolding in a heterogeneous group, it is imperative that the cognitive distance is not that wide as low-achievers may not understand high-achievers' explanations (Mugny and Doise 1978; Webb and Palincsar 1996). Also for high-achievers the cognitive distance with group members may play a role. High-ability students tend to transform any task into a meaningful *learning* task (Rijlaarsdam and Couzijn 2000) as they regulate their learning much better (Schunk 1995) and are capable of setting themselves learning goals (rather than only performance goals) (Dweck 1999). Therefore, high-ability students should profit from both heterogeneous and homogeneous grouping [i.e., irrespective of the ability level of the other group member(s)].

And indeed, some studies show that high-achievers benefit from collaboration for performance irrespective of grouping method (Hooper and Hannafin 1988; Leonard 2001). However, in other studies homogeneously grouped high-ability students outperformed their heterogeneously grouped counterparts (Fuchs et al. 1998; Webb et al. 2002). This might be explained by the distance range with the low-achiever in the heterogeneous groups. Webb (1982) for example shows that the frequency of explaining is positively related to achievement. It stands to reason that if the cognitive distance between high- and low-achiever is too large, explanations may not be solicited and/or offered (Fuchs et al. 1998; Mugny and Doise 1978). As a result, high-achievers may not profit as much as they should from (too) heterogeneous grouping.

Another factor that explains the effect of ability grouping next to group distance is task complexity (Fuchs et al. 1998; Webb et al. 2002). For ill-structured and open-ended tasks, it may be more productive for high-achievers to work with fellow high-achievers in a dyad with a relatively narrow range of ability. Fuchs et al. (1998) reason that for complex mathematics tasks in the pairs with a wide-ranging group ability they studied, offering well-reasoned explanations to low-achieving group members may have been too difficult for the high-achievers. For routine type of tasks that require clear-cut strategies and answers, the distance between the group members may be less of an issue (Webb et al. 2002).

We may expect that the effect of task complexity and distance between group members could be mediated by instruction. Indeed, instruction can significantly reduce task complexity for students in all ability groups, also the ones traditionally found to be least successful (low-ability homogeneous groups) (Wiedmann et al. 2012). However, only a minority of studies investigated the interaction between instruction and ability grouping (Hooper and Hannafin 1988; Wiedmann et al. 2012).

In writing and revision, Storch and Aldosari (2013) studied the effect of (EFL) proficiency pairing on EFL undergraduates' L2 use and interaction patterns during collaborative writing. However, the effect of proficiency pairing on text quality was not included nor was the interaction between proficiency pairing and instruction. Some studies on either revision or collaborative writing, mostly with children, recommend heterogeneous ability grouping but do not provide empirical data (cf. Francis and McCutchen 1994). They advocate heterogeneous grouping because low-ability students may profit from modelling of effective strategies by high-ability students. That this recommendation may especially hold for collaborative revision is suggested in a recently published study by Patchan et al. (2013). Patchan et al. studied the effect of ability level on the feedback undergraduates provide on each other's writing via the web-based reciprocal anonymous peer review system SWoRD (Cho and Schunn 2007). Results showed that low-ability writers "benefited more from the high-ability reviewers" (p. 397) as they received more comments focusing on substance issues from high-ability reviewers "which led to more feedback being implemented from the high-ability reviewers" (p. 397–398).

Studies that implement heterogeneous groups as an instructional method in collaborative writing are Piagetian and Vygotskian inspired instructional methods such as Paired Writing (Topping et al. 2000) and Peer Assisted Learning Strategies (PALS) (Fuchs et al. 1997; Saddler and Graham 2005) in which detailed procedures structure interaction and collaborative process writing in cross-ability pairs. In Paired Writing the more proficient writer—the tutor or Helper—guides the less proficient tutee or Writer during the various stages of the writing process with the aid of a flowchart outlining questions to arrive at a joint writing product (Sutherland and Topping 1999). These studies show that heterogeneous groups outperform students learning individually with respect to text quality (Sutherland and Topping 1999; Yarrow and Topping 2001).

One study (Saddler and Graham 2005) investigated the effect of sentence combining activities on 4th-grade PALS groups. The authors showed that pairs with a weaker and stronger student who alternatively acted as tutor and tutee in a sentence combining condition outperformed their counterparts in the comparison condition, a grammar instruction condition, for sentence combining and overall story writing quality.

One of the few studies in which the researchers next to cross-ability dyads (as in Saddler and Graham 2005) also investigate the effect of the instructional method, Paired Writing, on same-ability dyads is Sutherland and Topping (1999). In a small-scale pilot experiment, the researchers paired L1 8-year olds on the basis of teacher rankings of perceived pre-project writing ability, which, as the authors remarked, did not always correspond to initial pretest ability. The results show that Paired Writing is effective for both the less able Writers and their more able Helpers in the cross-ability fixed-role dyads. Even though same-ability dyads with reciprocating roles significantly improved their writing when writing collaboratively, the effect disappeared when students wrote individually.

Apart from Sutherland and Topping (1999), no studies in writing and revision research investigate the interaction between instructional methods and ability grouping factors. Additionally, the interaction for L1 or L2 learners in higher education is not explored. Nevertheless, as the effect of instruction may vary for different types of learners (Tobias 2010a, b), also in writing (cf. Galbraith 1992; Kieft et al. 2008), it is worth investigating the interaction for different pair configurations in L2 revision.

Revision instruction: modelling or learning-by-observation

Revision is an important, cognitively costly and complex process in writing (Chanquoy 2008; McCutchen 1996) which can occur at any point in the writing process (cf. Allal and Chanquoy 2004; Fitzgerald 1987): either prior to the actual writing or transcribing while planning ("pretextual revision" cf. Allal and Chanquoy 2004, p. 2); online, that is, while writing (e.g., revising a sentence or group of words just written) or when rereading a complete draft or part of the text ("deferred revision" cf. Allal and Chanquoy 2004, p. 2).

The revision process between more and less experienced writers and revisers differs significantly. More experienced revisers' task definition (i.e., revisers' knowledge, goals and scope of revision) is "more elaborate and effective" (Hayes 2004, p. 12). Secondly, more experienced revisers and writers (Fitzgerald 1992) do not only detect more problems, they also detect more problems on the global level of a text, so-called higher-order concerns (HOCs) dealing with content, structure and style, whereas less experienced writers mainly detect surface errors or so-called lower-order concerns (LOCs) (Fitzgerald 1992; Sommers 1980). This is in accordance with the developmental pattern of revision skill that follows a local to global developmental pattern (Berninger and Swanson 1994). Experts are also better at diagnosing the problems, have a larger and more sophisticated means-end table at their disposal, that is, more and more elaborate revision procedures and solutions (Faigley and Witte 1981; Fitzgerald 1992; Hayes 2004; Sommers 1980).

Even if in general writers revise more in their L2 than in their L1, regardless of their level of competency (Chenoweth and Hayes 2001; Silva 1993; Stevenson et al. 2006; Thorson 2000; Whalen and Ménard 1995), similar differences between experts and novices can be distinguished in L2 writing (New 1999; Stevenson et al. 2006). In fact, some researchers even speak of an intensification of the L1 revision pattern in L2 (Lindgren et al. 2008; Stevenson et al. 2006; Zamel 1983) or a "single system of revision across languages" (Hall 1990, p. 56) for advanced L2 writers.

In both L1 and L2 the subprocesses of revision (reading, error-detection and error-correction) are heavily governed by revisers' knowledge (Chenoweth and Hayes 2001; Hayes et al. 1987) e.g., about the audience (Chanquoy 2008) or about revision (Plumb et al. 1994; Wallace and Hayes 1991; Wallace et al. 1996); their goals, the scope of revision (*task definition* in Hayes et al. 1987; *task schema* in Hayes 1996) and working memory capacity (McCutchen 1996). Additionally, in the second or foreign language, L2 language proficiency plays a crucial role in distinguishing experts and novices. Manchón et al. (2009) show that more advanced L2 writers "engage four times more often in revision processes related to the elaboration and clarification of ideas, and to the solution of discourse and stylistic problems, than to compensatory language problems." (p. 113).

Consequently, if students do not revise or revise poorly either in L1 or in L2 it may be the result of a lack of knowledge (knowledge deficit cf. Chanquoy 2008) or because they fail to activate or retrieve knowledge (processing deficit) due to for example "an inadequate perception of the task" (Chanquoy 2008, p. 88) or "lack of (coordination needed in the) procedures needed for revision" (Chanquoy 2008, p. 88). Effective instructional methods to improve students' revision skill such as direct instruction in evaluative criteria (Hillocks 1986), structured peer interaction and reciprocal peer revision (MacArthur et al. 1991; Stoddard and MacArthur 1993) and forms of procedural facilitation (Bereiter and Scardamalia 1987; Chanquoy 2008) are thus targeted at either the knowledge deficit by increasing revisers' knowledge, the processing deficit by reducing the revision load or both.

Studies by Zimmerman and Kitsantas (2002) and Van Steendam et al. (2010) show that short-term revision strategy instruction via modelling (from the instructor's point of view) or observation (from the learner's point of view) is effective to improve both L1 and L2 undergraduates' revision skills respectively. In both studies modelling involves a short-term, explicit instruction of a strategy which is verbalized and demonstrated by models. The two studies illustrate the effectiveness of the first two stages of Schunk and Zimmerman's (1997) Social Model of Sequential Skill Acquisition according to which observation (Step 1) is followed consecutively by emulation in which students imitate the observed skill or strategy (Step 2). In the 4-step model, emulation is followed by self-directed practice to automatize the skill until learners can regulate their own performance.

Observing writers and/or revisers demonstrate their writing processes prior to executing the writing and/or revision task enables students to cope better with the double agenda of learning-to-write tasks. In learning-to-write tasks students are not only expected to write a qualitatively good text in a certain amount of time (task-execution agenda) but also expected to learn about and from writing (learning agenda) and to simultaneously "balance executional and monitoring processes for each of the agendas" (Rijlaarsdam and Couzijn 2000, p. 167). Consequently, an exclusive focus on the learning process first by observing models prior to task-execution may not only considerably reduce cognitive load (Rijlaarsdam and Couzijn 2000) also during subsequent task-execution because of "the alternation of executing and evaluating activities" (Rijlaarsdam and Couzijn 2000, p. 167) but may also—as a result of this alternation—result in more learning for both lower- and higher-ability students.

Especially for L2 students who, when revising other students' writing, tend to focus predominantly on surface-level errors (Fitzgerald 1987; Stevenson et al. 2006), a learning environment that triggers attention and reflection may be key (Bandura 1986; Rijlaarsdam and Couzijn 2000). For some L2 learners, this predominant focus on LOCs might be explained by a lower proficiency level in the L2. For others it may be due to the general abundance of LOCs in L2 texts which would distract from the HOCs. As in observational learning via instructional video also attentional processes are being triggered, observation may be especially productive for the acquisition of a global revision strategy by L2 learners, both poor and more proficient L2 revisers, following the adage "those who notice most, learn most" (Schmidt 1990, p. 144). The prevailing focus on form instead of on content by L2 revisers may also be the product of a long tradition of a focus on form in language classes (Lindgren et al. 2008). That is why especially L2 revisers in higher education might benefit from relatively short-term instruction via observation (to change their task representation). These students might have the required L2 language proficiency, but, as a result of a form-focused language education, may "define the task of revision [first and foremost] as making only local changes" (Wallace et al. 1996).

Empirical evidence illustrates that observation is more effective for all types of L1 secondary-school students than a more traditional task-execution or "learning-by-doing" condition (Couzijn 1999) in which learners get presentational instruction without teacher or peer modelling and then have to practise (Braaksma 2002; Rijlaarsdam et al. 2005).

Also for the complex problem-solving activity of collaborative revision in L2 learning-by-observation proves to be an effective instructional method (Van Steendam et al. 2010). When revising collaboratively students need to perform a number of cognitive activities simultaneously. On the one hand, there is the revision process in the foreign language which requires the detection, diagnosis and correction of textual problems both on the higher and lower levels of a text. On the other hand, there is the communication as students have to discuss and negotiate solutions with their pair members. Additionally, collaborative revision of higher-order textual problems is intrinsically complex because multiple solutions and strategies to deal with HOCs can be discerned, which are not guaranteed successful, that is, which do not necessarily result in a better text. Van Steendam et al. (2010) show that also undergraduate pairs revising collaboratively benefit more from observation than from practising for L2 revision quality. However, no studies investigate what the effect of ability grouping is when L2 pairs are instructed via modelling compared to a more traditional practising condition. Nevertheless, as drawing on the theoretical framework of observational learning in writing, learning-by-observation is expected to benefit both poor and more proficient writers and revisers, it stands to reason that both less able and more able learners in different ability pairs will benefit from the instruction.

This study

This study tests the effects of two pair composition factors and their interaction in two instructional settings on two dependent variables: revision and writing. The two instructional modi are (1) modelling or learning-by-observation and (2) practising or learning-by-doing. In the Observation condition students learn to revise and to write by observing peer models applying a revision strategy, while in the Practising condition students learn by practising the strategy themselves. Contrary to prior studies, ability grouping is operationalized by two continuous indices instead of a single dichotomous index: the proficiency level of individual pair members and the distance in proficiency between the pair members (heterogeneity). The interaction between these indices will be used as third characteristic of pair composition.

Research questions

Central to this study is the question whether the effect of ability grouping factors varies for two instructional methods (modelling or learning-by-observation and practising) in L2 collaborative revision. We will answer this question for two dependent variables, Revision Quality (learning) and Writing Quality (transfer).

Working hypotheses

As collaborative revision is a complex problem-solving task, following Fuchs et al. (1998) we expect that in the traditional practising condition initially stronger writers and revisers will benefit from homogeneous ability grouping (Hypothesis 1). The weaker students will profit more from the support of a more proficient writer/reviser in a heterogeneous ability group (Hypothesis 2).

Learning-by-Observation, on the other hand, is hypothesized to reduce task complexity for both weaker and stronger writers and revisers in heterogeneous and homogeneous dyads to the extent that they benefit from all ability groupings and that group composition does not play a role in this instructional mode (Hypothesis 3).

Method

Participants

The present study presents both a secondary and a supplementary analysis on data taken from another study (Van Steendam et al. 2010). Participants in the study are 160 Business freshmen attending an EFL course (42 % female) at a Belgian university. All students are first-year university students born in the same year. Mean age is 18.5 years ($SD = 3.17$ months). They are all native speakers of Dutch.

The participants belong to six class groups that were randomly (= alphabetically) composed by the university administration at the beginning of the academic year taking comparable class sizes into account (i.e., students with surnames from a to d were assigned to class group 1; class group 2 included students whose surnames started with e to h and so on; class 3^{i-l}; class 4^{m-p}; class 5^{q-t} and class 6^{u-z}). Each condition consists of three randomly-assigned classes.

The distribution of male and female students is not significantly different over conditions: $\chi^2(1, N = 160) = 0.59$, $p = 0.466$. Scores on the Oxford computer-based Quick Placement Test (QPT; Oxford University Press 2001) show that the majority of the students have an upper-intermediate to advanced EFL level: 10.4 % of the participants are lower-intermediate EFL users; 33.3 % of the participants have an upper-intermediate level of English; 32.6 % an advanced level and 23.6 % a very advanced level. The distribution of the different EFL proficiency levels is not significantly different over the two conditions: $\chi^2(3, N = 160) = 2.28$, $p = 0.517$. Table 2 provides more information about the pretest measures for the participants in the different conditions.

Design and procedure

We implemented a pretest—posttest design and administered pre- and posttests for two variables: revision and writing (cf. Table 1 Overview of Procedure). The intervention between pre- and posttests consisted of two phases: Instruction and Emulation in collaborative revision. Prior to the Instruction phase students were randomly assigned to dyads (hence: mixed- and same-gender dyads), which remained the same for the Emulation phase. The experimental intervention took place in the Instruction phase.

Instruction

In the instruction phase dyads were instructed in a strategy to revise content and structure of a text. The revision strategy itself consisted of six consecutive steps. Guiding questions helped students reflect on the content and structure of a letter of application and to detect

Table 1 Overview of procedure

Week 1: Pretest (individual)	Week 5: Intervention (dyad)	Week 6: Posttest (individual)
Writing → Revision	Instruction (- Observation, - Practising) → Emulation (- Revision)	Writing → Revision

HOCs. The strategy steps successively homed in on the content (Step 1) and structure of the letter as a whole [e.g., Did the writer include paragraphs? (Step 2); Are the paragraphs in a logical order? (Step 3)] and of every single paragraph separately [e.g., Does every paragraph have one main idea? (Step 4); Does every paragraph have a topic sentence? (Step 5) and Is every paragraph convincing? (Step 5)]. Every question representing a specific step was accompanied by an explanation. For every negative answer to a question, revision tips were suggested. The 'global' (i.e., as it deals with higher-order rather than local problems) revision strategy is genre-specific but at the same time has a high degree of transferability to other genres because it is a 'global' strategy to deal with structural and content problems which may appear in any type or genre of writing (e.g., a letter of application as well as an essay may lack unified paragraphs). The dyads were instructed in the revision strategy by means of two different instructional strategies corresponding to two conditions: a Practising condition and an Observation condition.

Practising condition (called PD for Practising in Dyad). In the PD condition, students received a handout with a letter of application for a university summer school. The letter contained structural and content problems (e.g., text without paragraphs). In randomly composed dyads students were asked to apply the revision strategy to the letter in consequent steps (step-by-step). Students first read the letter as a whole, after which they read the explications of a specific step in the handout. After application of the particular step for which they were provided with an explicit time indication, students were given a model answer on paper. This procedure was repeated for the six consecutive strategy steps. Students were told practising the strategy could enable them to revise and write a similar application letter. No further instructions on how dyad members had to interact or carry out the step-by-step strategy were provided to resemble authentic and naturalistic collaborative revision.

Observation condition (OD for Observation in Dyad). Students observed a pair of peer mastery models on CD-ROM demonstrate the revision strategy on the same letter that was the task set in the Practising condition. In dyadic interaction, the two students in the video demonstrated their application of the steps in the revision strategy and the detection and revision of HOCs. The changes the peer models made were the model answers provided in the PD condition. Students in the Observation condition were told they would watch two peers discuss and revise the structure and content of an application letter with the aid of a revision strategy. They were asked to pay close attention to the video to enable them to revise and write a similar letter of application. No further instructions were provided to the students in the Observation condition (cf. naturalistic and authentic situation).

The language of interaction in both conditions was Dutch, students' L1. Time-on-task in both conditions was kept equal. The time provided to dyads in the Practising condition to apply a strategy step was the same as the time the two models in the Observation condition spent on that step. To control for instructor effects, three experienced instructors were randomly assigned to both conditions and paper handouts with instructions, clarifications and model answers were used. As a result, every instructor was involved in both conditions and merely acted as a time-keeper.

Emulation phase

In both conditions an emulation session followed the instruction session. In this emulation session students applied the revision strategy and criteria in dyads to a peer's letter of application for a summer school on a topic having to do with students' field of study. The letter contained HOCs which students were expected to detect and revise with the help of the instructed strategy. Students used a procedural facilitator to keep check of the strategy steps (scaffolded dyadic practising), a cardboard sheet with an outline of the steps taught in the Instruction phase. Students could also view the strategy tool online.

All students worked on a computer, using Word Reviewing to make changes. Having both students work on the same computer to encourage them to pool their resources facilitated collaboration. The task, highly complex and open-ended as far as revision possibilities are concerned, requires joint problem-solving and interaction to enhance learning (Leonard 2001). To facilitate interaction, students were allowed to use their L1 during collaboration (De Guerrero and Villamil 2000).

Test measures

As Table 1 illustrates, students were administered three revision tests and two writing tests. Revision tests were an individual pretest, an emulation session test, part of the intervention (scaffolded dyadic revision), and an individual posttest (independent, i.e., unscaffolded, individual revision). Writing tests were an individual pretest and an individual posttest. All tests were done on a computer using Microsoft Word (i.e., for both revision and writing tests students used a PC). For all tests (pretests, emulation session test and posttests), students were given 1 h. As it was not possible to selectively administer a specific test version to a specific individual or dyad electronically, administering different versions for all the tests, even though preferable for generalization purposes and to avoid task effects (Schoonen 2012; Van den Bergh 1990), was not ecologically valid. For the posttest revision, however, a balanced design was used, that is, half of the students revised posttest A and half of the students revised version B. All tests and test prompts were vetted by Business English teachers for appropriate length, suitability and level of difficulty.

Revision tests

All revision tests consist of the same type of letter, a letter of application for a place on a summer school on an economics-related topic in which writers also enquire about accommodation and available scholarships. The authentic letters, not written by any of the participants, contain content and structural errors typical of the genre by the target group such as more than one main idea in a paragraph or an undeveloped paragraph (pretest: 10 HOCs; emulation and posttest: each 20 HOCs). The errors are not highlighted. For reasons of comparability and generalizability and to reduce error variance, none of the letters used were written by any of the students participating in the study. This had the added benefit that the required critical distance to identify discrepancies was created (Holliway and McCutchen 2004) and that students were not faced with peer pressure. Studies by Bartlett (1982) and Daneman and Stainton (1993) for example show that it is easier to notice word errors and ambiguous information in somebody else's writing than in one's own as a result of a required detachment from the text (Murray 1978; Rijlaarsdam et al. 2004).

For the three tests students were asked to revise on the structural and content level in English [i.e., off-line deep (as opposed to surface) revision, deferred revision]. For each error or target item students revised correctly, they received a score 1, and a zero when the item was not revised or not revised correctly. Revision Quality is indicated by the sum score of all the items correctly revised (Cronbach's alpha for the pretest is .70, for the emulation session test .89 and for the posttest revision .75). The pretest Revision Quality measure will subsequently referred to as initial (Individual) Revision Proficiency (cf. "Pair composition variables" section).

No further classification in terms of the different Actions that students performed when revising (e.g., additions, deletions, reorderings) or on the basis of Domain of revision (sentence, paragraph) was made (Stevenson et al. 2006) as in the instruction, even though different correct transformation possibilities in terms of Action or Domain were shown, no quality distinction was made between them. The Revision Quality score thus reflected the study's aim to teach students to revise HOCs correctly.

One rater scored all tests in a random order. A second rater independently scored 12.5 % of all the tests in a random order ($n = 60$). A scoring guide with benchmark examples for different correct transformation possibilities was provided. Reliability between the two sets of scores for Revision Quality was good ($\alpha = .89$, $p = .01$). Subsequently, disagreement between the discrepant scores was discussed and resolved.

Writing tests

At the beginning of the academic year students also wrote a letter of application for a summer school at an English-speaking university to prepare them for their Business studies. Five weeks later, after the instructional intervention, students wrote a second letter for a summer school on a topic related to their studies.

In the scoring system the traits assessed were the criteria presented to the participants in the instruction phase. As none of the existing text quality measures (e.g., Common European Framework of Reference, Council of Europe 2001) proved to be specific enough to capture the quality of the structural criteria instructed in the specific genre, we developed a set of scoring descriptors which were considered "to be salient in the context" (Hamp-Lyons 1995, p. 761). The scoring guide was thus based on the criteria for content and structure included in the revision strategy presented to students during the intervention and reflects the structural and content errors typical of the writing by the target group. Additionally, the criteria were also identified as important both for the quality of the specific genre and for writing quality in general and for the development of writing on the basis of reviewing textbooks, the literature on Business and Academic Writing, existing scoring descriptors for content and structure and by confering with expert-teachers of the genre and of the target audience. Students received a composite score of primary-trait scores for structure and content (out of 25) (Hamp-Lyons 1995; Lloyd-Jones 1977). These primary-trait scores were genre-specific and at the same time indicative of structural quality of the students' writing in general:

(1) a score for content (ranging from 0 to 10 points): the main question when scoring content is if authors succeeded in reaching their goal, that is, of being accepted for a place on a summer course. The answer was evaluated on the basis of the letter as a whole and every single paragraph. For every single paragraph (for example) raters verified if a paragraph contained sufficient and relevant information and was convincing enough to develop its topic and to reach its goal. Included in the content score is the degree to which students addressed the course's topic in their writing (e.g. in developing a motivation paragraph) instead of writing a sort of general, pass-key letter of application for any summer school on any topic. The score could thus also be labelled audience awareness.
(2) a score for structure (range 0–10 points): this score includes the evaluation of the structure of the letter as a whole (e.g., presence, number and order of paragraphs) and of every single paragraph (e.g., 'one paragraph is one idea' criterion);
(3) a score for topic sentences (5 points) in which both the presence and suitability of the topic sentences for every paragraph are evaluated.

Attenuated Pearson correlations between the three subscores are statistically significant and range from .55 to .65 (compare Lord and Novick 1968). Therefore the variables are assumed to be indicative of the same construct. The score will subsequently be referred to as Writing Quality for the posttest writing and as initial (Individual) Writing Proficiency (cf. "Pair composition variables" section). A large proportion of the anonimized typed writing assignments [$n = 159$, pretests ($n = 80$) and posttests ($n = 79$) mixed] were independently coded by two trained and experienced raters to compute inter-rater reliability and to verify the validity of the scoring rubric. The raters were experienced in assessing writing and in teaching English in general and (Business and Academic) writing in specific to EFL (Business) students. A scoring guide with benchmark examples of scores for the three dimensions was used to improve scoring reliability. The experienced raters were trained in a 2-h training session during which the scoring guide was explained, discussed in detail and applied to random sample copies. Raters convened a second time (a week later) after having scored a number of 25 randomly selected writing assignments. On the basis of a satisfactory interrater reliability ($r = .90$), raters individually scored the remaining writing assignments. Interrater reliability was fairly good (ranging from $r = .85$ to $r = .95$ with $p = .01$ for the different tests and traits).

Pair composition variables

On the basis of the two pretests, two indices were created to describe ability grouping: (1) level of proficiency (in revision and writing) of the individual pair member and (2) degree of heterogeneity in (writing/revision) proficiency within a dyad. The interaction term of these two indices allowed us to predict whether strong and weak learners benefited differently from pairs varying in heterogeneity. We constructed three continuous variables to predict for every initial writing and revision score the more effective intervention:

(1) Each member of a dyad received a continuous score for initial writing proficiency and initial revision proficiency. This measure is called the (initial) Individual (cf. subscript i) Proficiency measure (P_i): WP_i for writing and RP_i for revision.
(2) Every pair received a continuous score for the degree of heterogeneity in initial writing and revision proficiency between its members. This measure is the standard deviation of pretest performance within a dyad. It is an index for homogeneity/heterogeneity. We shall subsequently refer to this pair (cf. subscript p) variable as D_p for Distance or Deviation within a pair: RD_p for revision (Revision Distance in a pair) and WD_p for the writing measure (Writing Distance in a pair).
(3) Finally, for every dyad member the interaction between initial Individual Proficiency and Distance or Deviation within a pair was included. This interaction term is in fact a cross-level interaction. Combining the individual variable P_i and the degree of deviation in a pair D_p (in an interaction term cf. "Statistical Model" section) will allow us to make predictions for weaker and stronger writers and revisers in pairs that vary in heterogeneity.

Data analyses

Number of students

From the total number of 80 dyads or 160 students all of them participated in the emulation session. For the separate measures of the pretest revision (n = 6), pretest writing (n = 3), posttest revision (n = 7) and posttest writing (n = 4) we have a few missing cases (sometimes overlapping) which means that we sometimes analyzed data from 143 students or 71 dyads. However, since all students attended the emulation session, crucial to the design as this is where the intervention took place, and as multilevel models permit us to work with missing data (Hox 2002; Raudenbush and Bryk 2002), all students were included in the analyses.

Statistical model

Two multilevel models were analysed, using MLwiN 2.10 (Rasbash et al. 2009). In Model 1 for Revision Quality, for each of the two conditions one means for both posttests revision (emulation session test, that is, scaffolded dyadic revision, and the individual posttest, that is, unscaffolded, individual revision) was estimated next to three pair composition variables: (1) initial individual proficiency in revision (RP_i), (2) the standard deviation of initial revision proficiency within a dyad (Revision Distance in a pair RD_p) and (3) the interaction between RP_i and RD_p. Hence, in total 8 fixed parameters (2 fixed parameters for the two conditions * 4 fixed parameters, that is, intercept and three pair composition variables) and three variance components are estimated. As revision tests (level 1) were nested in individuals (level 2) that were nested in pairs (level 3), a multivariate multilevel fixed occasion model with three levels was used (Goldstein 2003).

In Model 2 for Writing Quality, for each of the two conditions the intercept was estimated next to 3 pair composition variables: (1) initial Individual Writing Proficiency (WP_i), (2) the degree in deviation in a pair for initial Writing Proficiency (Writing Distance in a pair WD_p) and (3) the interaction of these two pair composition terms. Hence in total 2 (for the two conditions) * 4 fixed parameters (intercept, three pair composition variables) and two variances were estimated. This model is a univariate model in which individuals (level 1) are nested in pairs (level 2) (Quené and Van den Bergh 2004).

Additionally, in every single condition statistical significance of a variable is tested by computing the Wald Test (i.e., the ratio of the estimate and its standard error is standard normal distributed).

Results

Preliminary analyses

Means and standard deviations for pretest and posttest measures can be found in Table 2. No statistically significant differences in pretest measures were found between the two conditions for QPT ($\chi^2(1, N = 157) = 2.35, p = .13$) nor for the initial Individual Writing Proficiency measure ($\chi^2(1, N = 157) = 0.235, p = .63$) or for the initial Individual Revision Proficiency measure ($\chi^2(1, N = 154) = 3.5, p = .15$). Additionally, for the posttest revision Revision Quality no statistically significant task differences were observed between the two versions ($\chi^2(1, N = 153) = 1.7, p = .19$).

Table 2 Means and Standard Deviations (in Brackets) for Conditions for Pretest and Post-test Measures

	Score out of	Conditions: Observation (n = 88)	Practising (n = 72)
Oxford Quick Placement Test	Pre /100	72.33 (9.94)	69.83 (9.96)
Writing Quality[a]	Pre	0.11 (1.04)	−0.14 (0.88)
	Post	0.15 (1.04)	−0.35 (0.88)
Revision Quality[a]	Pre	0.42 (0.99)	−0.23 (0.5)
	Post	0.17 (1.03)	−0.29 (0.92)

[a] Standardized scores

Learning effect: revision quality

In the Practising condition, no statistically significant effects were found for Individual Revision Proficiency RP_i ($\beta = -0.068$ (SE = 0.167), $p = .68$) nor for Revision Distance in a pair RD_p ($\beta = 0.179$ (0.417), $p = .67$) cf. Table 3 for regression coefficients and standard errors for pair composition variables (Fixed Coefficients). However, the interaction between these factors was significant ($\beta = -0.575$ (0.260), $p = .02$). Thus within this instructional modus the learning effect on revision depends on the combination of individual students' Revision Proficiency and the degree of heterogeneity in Revision Proficiency within the dyad students belongs to. Figure 1 illustrates that the better individual revisers scored on the pretest revision, the less they benefited from a heterogeneous pair in PD, while the lower students scored for initial Revision Proficiency, the more they prospered in a heterogeneous dyad. The variance explained by the interaction of both RP_i and RD_p in PD is 6 %, which corresponds to a medium effect size (cf. Cohen 1977).

In the Observation condition (OD in Fig. 1) we found no statistically significant effects, neither for initial Individual Revision Proficiency RP_i ($\beta = 0.215$ (0.776), $p = .78$), nor for Revision Distance in a pair RD_p ($\beta = 0.116$ (0.170), $p = .49$) or the interaction RP_{i*} RD_p ($\beta = -0.041$ (0.110), $p = .71$).

Table 3 provides the variances attributed to the different levels in the multilevel fixed occasion model (Random Coefficients).

Transfer effect: writing quality

In the Practising condition we observed a statistically significant effect for initial Individual Writing Proficiency ($\beta = 0.409$ (0.176), $p = .02$). The higher individual writers scored on the pretest writing, the more they prospered in these instructional modi (see Fig. 2).

We found the same result in the Observation condition ($\beta = 0.414$ (0.152), $p = .001$) cf. Table 4 for regression coefficients and standard errors. The better individual writers are, the more they benefit in this instructional modus. The variance explained by students' initial Writing Proficiency in the Observation condition is 9.4 % corresponding to a medium effect size (Cohen 1977).

Table 3 Parameter Estimates (Fixed and Random) for Revision Quality

	Fixed Regression Coefficients (Standard Errors)	
	Conditions	
	Observation	Practising
Initial Individual Revision Proficiency	0.215	−0.068
	(0.776)	(0.167)
Revision Distance in a pair	0.116	0.179
	(0.170)	(0.417)
Interaction	−0.041	−0.575*
Initial Revision Proficiency* Revision Distance in a pair	(0.110)	(0.260)
		Random Coefficients
Variance between pairs		0.249 (0.068)*
Variance between individuals		0.010 (0.052)*
Variance within individuals		0.726 (0.065)*

* $p < .05$

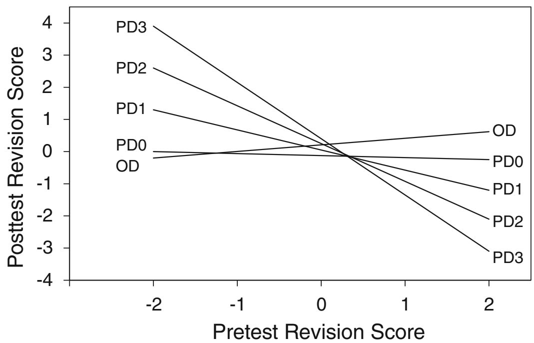

Fig. 1 The *vertical lines* (PD 0 to PD 3) represent the different standard deviations in PD ranging from D = 0 for a completely homogeneous dyad to D = 3 for the highest between-individual difference within a pair. The higher individual writers score on the pretest revision (cf. standardized z-scores on the X-axis), the more they benefit from a homogeneous pair in PD for the posttest revision (cf. standardized z-scores on the Y-axis)

We found no statistically significant effects for Writing Distance in a pair WD_p in either the Practising ($\beta = 0.041$ (0.089), $p = .65$) or the Observation condition ($\beta = -0.008$ (0.068), $p = .91$).

However, in Practising the interaction between WP_i and WD_p is statistically significant ($\beta = -0.138\,(0.067), p = .03$). This means that the transfer effect in this instructional modus depends on *both* individual students' initial Writing Proficiency and the degree of distance in Writing Proficiency. Figure 2 illustrates that the wider the distance in initial writing proficiency between the pair members in PD, the more the less proficient writers benefit from a

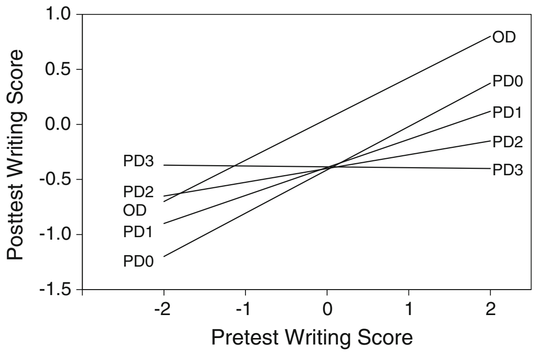

Fig. 2 The different *horizontal lines* (PD0 to PD3) represent standard deviations ranging from 0 for completely homogeneous dyads (PD = 0) to 1, 2 and 3 for increasingly more heterogeneous ability dyads

Table 4 Parameter Estimates (Fixed and Random) for Writing Quality

	Fixed Regression Coefficients (Standard Errors)	
	Conditions	
	Observation	Practising
Initial Writing Proficiency	0.414*	0.409*
	(0.152)	(0.176)
Writing Distance in a pair	−0.008	0.041
	(0.068)	(0.089)
Interaction	−0.054	−0.138*
Initial Writing Proficiency* Writing Distance in a pair	(0.059)	(0.067)
		Random Coefficients
Variance between pairs		0.399 (0.080)*
Variance within pairs and between individuals		0.559 (0.041)*

* $p < .05$

heterogeneous pair as opposed to the more proficient writers. The variance explained by the interaction of both WP_i and WS_p in Practising is 5 % (=medium effect size).

In the Observation condition no statistically significant effects were found for the interaction of initial Individual Writing Proficiency and Writing Distance in a pair ($\beta = -0.054$ (0.059), $p = .36$).

Table 4 presents the variances attributed to the different levels in the multilevel fixed occasion model (Random Coefficients).

Discussion

In the present study we tested the effect of pair composition factors on learning and transfer results in two instructional modi for collaborative revision in L2, to test the compensating effect of instructional support. We set two indicators and their interaction to define ability grouping, three *continuous* scores on the basis of pretests paralleling skills targeted during instruction and in the outcome measure. The fact that, apart from students' individual proficiency, we also included the distance in initial proficiency between pair members allowed us to test the effect of an increasingly more or less wide-ranging proficiency distance on both low- and high-achievers.

The results show that the effect of pair composition depends on instructional strategies. In the *Practising condition* the more proficient L2 revisers are, the less they benefit for *Revision Quality* from working together with a less proficient reviser in a heterogeneous dyad. However, less skilled revisers do profit from collaborating with a more skilled reviser. These results are corroborated for L2 writers for *Writing Quality*. In this study a global revision strategy for content and structure was instructed with the two-fold aim of improving L2 students' higher-order revision skills and consequently having them prioritize content organisation when writing, resulting in better texts. That there is a carry-over effect from revision to writing of the same type of texts, may supply evidence for the 'learning-by-reviewing' hypothesis put forward by Cho and MacArthur (2011) which states that students improve their own writing as a result of reviewing others'.

The finding that less proficient revisers and writers benefit from the support of a stronger reviser or writer, even if quite frequently at the latter's expense, confirms Hypothesis 2 and is in line with prior research on ability grouping in collaborative learning (Hooper and Hannafin 1991; Webb et al. 2002; Wiedmann et al. 2012).

That vice versa strong writers are preferably grouped with another high-ability writer in a homogeneous dyad (confirmation of Hypothesis 1) is consistent with studies by Webb et al. (2002) and Fuchs et al. (1998). Contrary to what Webb et al. (2002) found for less complex tasks in science education, for the complex task of collaborative revision and subsequent application of a content/structure first strategy to a new draft, the level of the less proficient pair member in heterogeneous pairs does play a role for the high-ability students. The result that the wider the distance in skill with the less proficient learner, the less the initially more proficient learners benefit from the collaboration, seems support for a strong relation between ability grouping and task complexity. At the same time, the results suggest that the effect of ability grouping and task complexity cannot be seen separately from instruction.

As in the *Observation condition* no statistically significant effects on *Revision Quality* are found either for Individual Revision Proficiency or for degree of heterogeneity in Revision Proficiency in a Dyad (Revision Distance in a pair), we can conclude that revisers in any type of dyad benefit from the condition. This result illustrates that Observation in collaborative revision is a powerful instructional method as it overrides some of the individual differences which can be found in a more traditional practising condition. The instructional method thus seems to be effective in changing these EFL college students' task representation (i.e., having them revise a text for content and structure first) and in reducing task complexity for both lower- and higher-ability learners to the degree that both benefit from heterogeneous and homogeneous grouping (cf. Hypothesis 3). Future research could also include (indirect) measures of cognitive load (e.g., Paas et al. 2003) to confirm this. Also, when students are instructed in a global revision strategy via modelling, distance in proficiency between these learners does not seem to matter.

However, Hypothesis 3 is only partially confirmed as results for *Writing Quality*, the group-to-individual transfer variable, show that in Observation the more proficient writers benefit more from the condition than the less proficient writers, regardless of dyad composition. That the transfer effect from collaborative revision to subsequent individual writing is mediated by writing ability confirms previous research on instruction and ability grouping in statistics (Wiedmann et al. 2012).

That the stronger writers benefited more from the observation of a mastery peer revision model for writing than the weaker writers could first of all confirm that stronger students traditionally seem to outperform weaker writers. It can also be taken as support for the similarity hypothesis validated by Braaksma et al. (2002), which states that students initially benefit more from focusing on a model similar to them in performance. However, as in our study students were instructed in the fairly complex problem-solving activity of revising HOCs in a new genre, we hypothesized a mastery model, apart from being effective for the stronger learners, also to be less cognitively demanding and less confusing for the weak novice revisers to learn about the detection, diagnosis and revision of problems than a model showing a poor collaborative revision dyad (a so-called weak model cf. Braaksma 2002) or a coping model of a dyad first doing poorly but gradually improving its performance (Kitsantas et al. 2000). Research in the field of worked examples in science and mathematics education, in which learners are presented with expert models of cognitive problem-solving, also advocates an error-free expert model in the initial stages of skill acquisition and cognitive problem-solving and recommends postponing incorrect solutions to a later stage in problem-solving (Große and Renkl 2007).

For Revision Quality, observing a peer mastery model of collaborative revision proved to be beneficial to both weak and stronger revisers. However, for transfer to Writing Quality, less proficient writers did not seem to benefit as much as the more proficient writers. Perhaps writers in the weak dyads did not have the required frame of reference to fully transfer the revision strategy demonstrated by the mastery dyad to their own writing. Future research should experimentally test if weak low-ability students are better served by a weak model dyad for transfer to writing or by a sequence of mastery and weaker models. On the other hand, it should be noted that it is not because initially weaker writers did not benefit *to the same degree* as more proficient writers for Writing Quality that learning was not taking place.

In a study with Dutch L1 high-school students, Braaksma et al. (2004) found a significant impact of observation on students' writing processes: students' writing processes were more varied than in a direct writing condition. Consequently, it is highly likely that in the present study weak students in the observation condition changed their writing and revision process to a significant degree or evolved in their reflection on texts, as the results for Revision Quality may indicate, but did not translate this into qualitatively better texts than the stronger writers. Future research should include writing and revision process data to verify this and to offer an insight into students' actual strategy use and task representation.

Studies by Winters and Alexander (2011) and Barron (2003) in collaborative science and mathematics problem-solving with high-school students and 6-grade students respectively illustrate that interaction may have a significant impact on learning and transfer. That is why in further research quality and quantity of interaction (Hooper and Hannafin 1991; Villamil and De Guerrero 1996), learning mechanisms at play in interaction such as explaining, giving and receiving elaborated help (Webb 1982), instances of sociocognitive conflict (Fuchs et al. 1998) and of regulatory behaviour (Winters and Alexander 2011) should be studied to determine the full impact of interaction, instruction and group composition (Webb and Palincsar 1996).

Despite the fact that we need replication studies, long-term and with different tasks, to generalize these findings, the strengths of this study are first of all that it confirms that short-term revision strategy instruction in collaboration may result in rapid significant improvement in writing quality for college-level students. Wallace and Hayes for example illustrated that a mere 8-min instructional prompt to revise more globally resulted in a change in entry-level college writing students' task schemata for revision tasks and a consequent improvement in writing and revision performance both of their own (Wallace et al. 1996) and other students' texts (Wallace and Hayes 1991). The present study shows that short-term revision instruction via observation in collaborative revision of other students' writing is an effective instructional method to improve undergraduates' revision skills in a foreign language irrespective of grouping method (heterogeneous and homogeneous). At the same time, the study also illustrates that what 'works' for more proficient revisers and writers in terms of grouping in more traditional instructional conditions is not always equally effective for poor revisers and writers and vice versa.

In this respect, the study attests to the importance of aptitude treatment interaction (ATI) research (Cronbach and Snow 1977; Tobias 2010a, b), which starts from the premise that the effect of instructional strategies may be different for different types of learners and that instructional treatments should be adapted to the aptitude or proficiency of individual learners to result in better achievement and learning, also in writing (Galbraith 1992; Braaksma 2002). Not only does the present study provide support for ATI research but also for the educational practice of matching instructional strategies and students. As such, the results contribute significantly to our understanding of effective strategy instruction for collaborative revision in L2 for different types of foreign language revisers and writers in heterogeneous and homogeneous pairs. Additionally, the study provides us with a first insight into the complex interplay between pair composition and observation for the fairly complex task of revising collaboratively in L2.

Implications for educational practice seem to be that teachers implementing collaborative revision in L2 without modelling should be aware of the complex trade-off of mixed-ability groups for weak and strong writers and revisers. Observation, on the other hand, seems to override this precarious balance for weak and strong revisers in heterogeneous groups and emerges as an effective instructional method for collaborative revision in L2 for both heterogeneous and homogeneous ability dyads.

References

Allal, L., & Chanquoy, L. (2004). Introduction: Revision revisited. In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes* (Vol. 13, pp. 1–7). Dordrecht: Kluwer Academic Publishers.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*, 307–359. doi:10.1207/S15327809JLS1203_1.

Bartlett, E. J. (1982). Learning to revise: Some component processes. In M. Nystrand (Ed.), *What writers know: The language, process, and structure of of written discourse* (pp. 345–363). New York: Academic Press.

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

Berninger, V. W., & Swanson, H. L. (1994). Modification of the Hayes and Flower model to explain beginning and developing writing. In E. Butterfield (Ed.), *Advances in cognition and educational practice. Children's writing: Toward a process theory of development of skilled writing* (Vol. 2, pp. 57–82). Greenwich, CT: JAI Press.

Braaksma, M. (2002). Observational learning in argumentative writing (Doctoral thesis, University of Amsterdam, The Netherlands).

Braaksma, M., Rijlaarsdam, G., & Van den Bergh, H. (2002). Observational learning and the effects of model–observer similarity. *Journal of Educational Psychology, 94*(2), 405–415. doi:10.1037//0022-0663.94.

Braaksma, M., Rijlaarsdam, G., van den Bergh, H., & Hout Wolters, B. A. M. (2004). Observational learning and its effects on the orchestration of the writing process. *Cognition & Instruction, 22*, 1–36. doi:10.1207/s1532690Xci2201_1.

Chanquoy, L. (2008). Revision processes. In R. Beard, D. Myhill, J. Riley, & M. Nystrand (Eds.), *The Sage handbook of writing development* (pp. 80–97). Thousand Oaks, CA: Sage Publications Ltd.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing. Generating text in L1 and L2. *Written Communication, 18*(1), 80–98.

Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of Educational Psychology, 103*(1), 73–84. doi:10.1037/a0021950.

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education, 48*(3), 409–426. doi:10.1016/j.compedu.2005.02.004.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev ed.). New York: Academic Press.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg: Cambridge University Press.

Couzijn, M. (1999). Learning to write by observation of writing and reading processes: Effects on learning and transfer. *Learning and Instruction, 9*, 109–142. doi:10.1016/S0959-4752(98)00040-1.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods. A handbook for research on interactions*. New York: Irvington.

Daiute, C., & Dalton, B. (1993). Collaboration between children learning to write: Can novices be masters? *Cognition and Instruction, 10*(4), 281–333.

Daneman, M., & Stainton, M. (1993). The generation effect in reading and proofreading. *Reading and writing, 5*(3), 297–313. doi:10.1007/BF01027393.

De Guerrero, M. C. M., & Villamil, O. S. (2000). Activating the ZPD: Mutual scaffolding in L2 peer revision. *Modern Language Journal, 84*(1), 51–68. doi:10.1111/0026-7902.00052.

De Lisi, R., & Golbeck, S. L. (1999). Implications of Piagetian theory for peer learning. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 3–38). Mahwah, NJ: Lawrence Erlbaum Associates.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Taylor and Francis.

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

Fitzgerald, J. (1987). Research on revision in writing. *Review of Educational Research, 37*(5), 481–506.

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

Francis, M., & McCutchen, D. (1994, April 4–8). *Strategy differences in revising between skilled and less-skilled writers.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Karns, K. (1998). High-achieving students' interactions and performance on complex mathematical tasks as a function of homogeneous and heterogeneous pairings. *American Educational Research Journal, 35*(2), 227–267. doi:10.3102/00028312035002227.

Fuchs, D., Fuchs, L., Mathes, P., & Simmons, D. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Association, 34*, 174–206. doi:10.3102/00028312034001174.

Galbraith, D. (1992). Conditions for discovery through writing. *Instructional Science, 21*, 45–73. doi:10.1007/BF00119655.

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

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction, 17*(6), 612–634. doi:10.1016/j.learninstruc.2007.09.008.

Hall, C. (1990). Managing the complexity of revision across languages. *TESOL Quarterly, 24*, 245–266.

Hamp-Lyons, L. (1995). Research on the rating process. Rating nonnative writing: the trouble with holistic scoring. *TESOL Quarterly, 29*, 759–762. doi:10.2307/3588173.

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. (2004). What triggers revision? In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes* (Vol. 13, pp. 9–20). Dordrecht: Kluwer Academic Publishers.

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

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

Hogan, D. M., & Tudge, J. R. H. (1999). Implications of Vygotsky's theory for peer learning. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 39–66). Mahwah, NJ: Lawrence Erlbaum Associates.

Holliway, D. R., & McCutchen, D. (2004). Audience perspective in young writers' composing and revising. Reading as the reader. In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes* (Vol. 13, pp. 87–101). Dordrecht: Kluwer Academic Publishers.

Hooper, S., & Hannafin, M. J. (1988). Cooperative CBI: The effects of heterogeneous versus homogeneous grouping on the learning of progressively complex concepts. *Journal of Educational Computing Research, 4*(4), 413–424.

Hooper, S., & Hannafin, M. J. (1991). The effects of group composition on achievement, interaction, and learning efficiency during computer-based cooperative instruction. *Educational Technology Research and Development, 39*(3), 27–40. doi:10.1007/BF02296436.

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

Kieft, M., Rijlaarsdam, G., & Van den Bergh, H. (2008). An aptitude-treatment interaction approach to writing-to-learn. *Learning and Instruction, 18*(4), 379–390. doi:10.1016/j.learninstruc.2007.07.004.

Kitsantas, A., Zimmerman, B. J., & Cleary, T. (2000). The role of observation and emulation in the development of athletic self-regulation. *Journal of Educational Psychology, 92*(4), 811–817. doi:10.1037//0022-0663.92.4.811Development.

Leonard, J. (2001). How group composition influenced the achievement of sixth-grade mathematics students. *Mathematical Thinking and Learning and Instruction, 3*(2), 175–200. doi:10.1080/10986065.2001.9679972.

Lindgren, E., Spelman Miller, K., & Sullivan, K. P. H. (2008). Development of fluency and revision in L1 and L2 writing in Swedish high school years eight and nine. *ITL—International Journal of Applied Linguistics, 156*, 133–151. doi:10.2143/ITL.156.0.2034428.

Lloyd-Jones, R. (1977). Primary-trait scoring. In C. R. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 33–68). Urbana, IL: National Council of Teachers of English.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

MacArthur, C., Schwartz, S., & Graham, S. (1991). Effects of a reciprocal peer revision strategy in special education classrooms. *Learning Disabilities Research and Practice, 6*(4), 201–210.

Manchón, R. M., Roca de Larios, J., & Murphy, L. (2009). The temporal dimension and problem-solving nature of foreign language composing processes. Implications for theory. In R. M. Manchón (Ed.), *Writing in foreign language contexts. Learning, teaching, and research* (pp. 102–129). Clevedon, UK: Multilingual Matters.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review, 8*(3), 299–325.

Min, H. T. (2005). Training students to become successful peer reviewers. *System, 33*(2), 293–308. doi:10.1016/j.system.2004.11.003.

Mugny, G., & Doise, W. (1978). Socio-cognitive conflict and structure of individual and collective performance. *European Journal of Social Psychology, 8*, 181–192.

Murray, D. M. (1978). Internal revision: A process of discovery. In C. R. Cooper & L. Odell (Eds.), *Research on composing: Points of departure* (pp. 85–103). Urbana, IL: National Council of Teachers of English.

New, E. (1999). Computer-aided writing in French as a foreign language: A qualitative and quantitative look at the process of revision. *The Modern Language Journal, 83*(1), 80–97. doi:10.1111/0026-7902.00007.

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

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63–71. doi:10.1207/S15326985EP3801_8.

Patchan, M. M., Hawk, B., Stevens, C. A., & Schunn, C. D. (2013). The effects of skill diversity on commenting and revisions. *Instructional Science, 41*(2), 381–405. doi:10.1007/s11251-012-9236-3.

Piaget, J. (1932). *The language and thought of the child* (2nd ed.). London: Routledge & Kegan Paul.

Plumb, C., Butterfield, E. C., Hacker, D. J., & Dunlosky, J. (1994). Error correction in text. Testing the processing deficit and knowledge deficit hypotheses. *Reading and Writing, An Interdisciplinary Journal, 6*(4), 347–360. doi:10.1007/BF01028848.

Quené, H., & Van den Bergh, H. (2004). On multilevel modeling of data from repeated measures: A tutorial. *Speech Communication, 43*, 103–121. doi:10.1016/j.specom.2004.02.004.

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

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Kieft, M., Broekkamp, H., et al. (2005). Psychology and the teaching of writing in 60 minutes. In P. Tomlinson, J. Dockrell, & P. Winne (Eds.), *BJEP Monograph Series II, No. 3: Pedagogy—Teaching for Learning* (pp. 127–153). Leicester: The British Psychological Society. doi:10.1348/000709905X62156.
Rijlaarsdam, G., & Couzijn, M. (2000). Writing and learning-to-write. A double challenge. In R. Simons, J. Van der Linden, & T. Duffy (Eds.), *New learning* (pp. 157–190). Dordrecht: Kluwer.
Rijlaarsdam, G., Couzijn, M., & Van den Bergh, H. (2004). The study of revision as a writing process and as a learning-to-write process. Two prospective research agendas. In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes* (Vol. 13, pp. 189–207). Dordrecht: Kluwer Academic Publishers.
Saddler, B., & Graham, S. (2005). The effects of peer-assisted sentence-combining instruction on the writing performance of more and less skilled young writers. *Journal of Educational Psychology, 97*(1), 43–54. doi:10.1037/0022-0663.97.1.43.
Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.
Schoonen, R. (2012). The validity and generalizability of writing scores: The effect of rater, task and language. In E. Van Steendam, M. Tillema, G. Rijlaarsdam, & H. Van den Bergh (Eds.), *Measuring Writing: Recent insights into Theory, Methodology and Practices* (Vol. 27, pp. 1–22). Leiden: Koninklijke Brill NV.
Schunk, D. H. (1995). Self-efficacy and education and instruction. In J. E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment: Theory, research, and application* (pp. 281–303). New York: Plenum Press.
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*(4), 657–677. doi:10.2307/3587400.
Sommers, N. (1980). Revision strategies of student writers and experienced writers. *College Composition and Communication, 31*, 378–387.
Stevenson, M., Schoonen, R., & De Glopper, K. (2006). Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing, 15*, 201–233. doi:10.1016/j.jslw.2006.06.002.
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*(1), 76–103.
Storch, N., & Aldosari, A. (2013). Pairing learners in pairwork activity. *Language Teaching Research, 17*(31), 31–48. doi:10.1177/1362168812457530.
Sutherland, J. A., & Topping, K. J. (1999). Collaborative creative writing in eight-year olds: Comparing cross-ability, fixed and same-ability reciprocal role pairing. *Journal of Research in Reading, 22*(2), 154–179. doi:10.1111/1467-9817.00080.
Thorson, H. (2000). Using the computer to compare foreign and native language writing processes: A statistical and case study approach. *Modern Language Journal, 84*(2), 155–170.
Tobias, S. (2010a). The expertise reversal effect and aptitude treatment interaction research (Commentary). *Instructional Science, 38*, 309–314. doi:10.1007/s11251-009-9103-z.
Tobias, S. (2010b). Aptitudes and instructional methods. In K. L. Rasmussen (Ed.), *Encyclopedia of research design* (pp. 38–40). New York: Sage Publications. doi:10.4135/9781412961288.
Topping, K., Nixon, J., Sutherland, J., & Yarrow, F. (2000). Paired writing: A framework for effective collaboration. *Reading (UKRA), 34*(2), 79–89. doi:10.1111/1467-9345.00139.
Van den Bergh, H. (1990). On the construct validity of multiple choice items for reading comprehension. *Applied Psychological Measurement, 14*(1), 1–12. doi:10.1177/014662169001400101.
Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction, 20*(4), 316–327. doi:10.1016/j.learninstruc.2009.08.009.
Villamil, O. S., & De Guerrero, M. C. M. (1996). Peer revision in the L2 classroom: Social-cognitive activities, mediating strategies, and aspects of social behaviour. *Journal of Second Language Writing, 5*(1), 51–75. doi:10.1016/S1060-3743(96)90015-6.
Villamil, O. S., & De Guerrero, M. C. M. (1998). Assessing the impact of peer revision on L2 writing. *Applied Linguistics, 19*(4), 491–514. doi:10.1093/applin/19.4.491.
Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wallace, D. L., & Hayes, J. R. (1991). Redefining revision for freshmen. *Research in the Teaching of English, 25*(1), 54–66.

Wallace, D. L., Hayes, J., Hatch, J., Miller, W., Moser, G., & Silk, C. (1996). Better revision in eight minutes? Prompting first-year college writers to revise globally. *Journal of Educational Psychology, 88*(4), 682–688.

Webb, N. M. (1980). A process-outcome analysis of learning in group and individual settings. *Educational Psychologist, 15*, 69–83. doi:10.1080/00461528009529217.

Webb, N. M. (1982). Peer interaction and learning in cooperative small groups. *Journal of Educational Psychology, 74*(5), 642–655. doi:10.1016/0883-0355(89)90014-1.

Webb, N. M., Nemer, K. M., Chizhik, A. W., & Sugrue, B. (1998). Equity issues in collaborative group assessment: Group composition and performance. *American Educational Research Journal, 35*(4), 607–651. doi:10.3102/00028312035004607.

Webb, N. M., Nemer, K. M., & Zuniga, S. (2002). Short circuits or superconductors? Effects of group composition on high-achieving students' science assessment performance. *American Educational Research Journal, 39*(4), 943–989. doi:10.3102/00028312039004943.

Webb, N. M., & Palincsar, A. S. (1996). Group processes in the classroom. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 841–873). New York: Macmillan.

Whalen, K., & Ménard, N. (1995). L1 and L2 writers' strategic and linguistic knowledge: A model of multiple-level discourse processing. *Language Learning, 45*(3), 381–418. doi:10.1111/j.1467-1770.1995.tb00447.x.

Wiedmann, M., Leach, R. C., Rummer, N., & Wiley, J. (2012). Does group composition affect learning by invention? *Instructional Science, 40*, 711–730. doi:10.1007/s11251-012-9204-y.

Winters, F. I., & Alexander, P. A. (2011). Peer collaboration: The relation of regulatory behaviors to learning with hypermedia. *Instructional Science, 39*, 407–427. doi:10.1007/s11251-010-9134-5.

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

Yarrow, F., & Topping, K. (2001). Collaborative writing: The effects of metacognitive prompting and structured peer interaction. *British Journal of Educational Psychology, 71*(2), 261–282. doi:10.1348/000709901158514.

Zamel, V. (1983). The composing process of advanced ESL students: Six case studies. *TESOL Quarterly, 17*(2), 165–187. doi:10.2307/3586647.

Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94*(4), 660–668. doi:10.1037/0022-0663.94.4.660.