Comprehensive corrective feedback on low and high proficiency writers
Examining attitudes and preferences

Marisela Bonilla López, Elke Van Steendam and Kris Buyse
KU Leuven

This study investigated the effects of comprehensive feedback on learners’ grammatical accuracy during text revision and in new writing tasks in light of proficiency level. It also sought to determine to what extent learners’ proficiency level plays a role in their feedback preferences and attitudes towards the feedback. The participants were 52 low proficiency and 39 high proficiency foreign language university learners, who were randomly assigned to a direct corrective feedback, a metalinguistic feedback with rule reminders, and a self-correction group. All learners wrote four compositions and completed a questionnaire after the treatment to elicit their attitudes towards the feedback and their feedback preferences. Results showed that the treatment effectively enhanced both low and high proficiency learners’ immediate grammatical accuracy and accuracy improvement. Also, a relation between proficiency level and learners’ attitudes towards the feedback as well as an association between proficiency level and learners’ feedback preferences were found.

Keywords: direct corrective feedback, feedback attitudes, feedback preferences, high proficiency learners, metalinguistic feedback, low proficiency learners

Introduction

Despite arguments against written corrective feedback (CF) or error correction (e.g., Semke, 1984; Truscott, 1999), it still is an essential component for both teachers and students in the second language (L2) writing class (Ferris, 2010). For this reason, if anything, Truscott (1996) ignited more research interest in error correction and its efficacy rather than a complete abandonment of the practice. Since then, various studies have provided evidence that demonstrates that corrective feedback (CF) increases learners’ grammatical accuracy during text revision (e.g.,
Ferris, 2006) and yields a learning effect in new writings (e.g., van Beuningen, De Jong, & Kuiken, 2012). Unfortunately, comparing WCF studies has proven difficult because the design, treatment, participants, and instructional contexts are overwhelmingly dissimilar (Hyland & Hyland, 2006).

Also, while CF studies from a second language acquisition (SLA) perspective (e.g., Bitchener & Knoch, 2010a, 2010b; Shintani & Ellis, 2013; Shintani, Ellis, & Suzuki, 2014) have advanced significantly in tackling design shortcomings of previous WCF studies (see Guénette, 2007 and Storch, 2010 for a review), intervening variables beyond the students’ written products tend to be overlooked in CF studies (Ferris, Liu, Sinha, & Senna, 2013; Storch, 2010). For this reason, echoing Bitchener and Ferris’ (2012) call for feedback studies that look into the interactional effect of CF with other variables, the present study intends to make a start in that direction by examining the effectiveness of comprehensive CF in conjunction with learners’ proficiency level. In addition, given the need for more evidence concerning the role of affective variables in feedback studies (Ferris & Roberts, 2001), this study will also look into learners’ feedback preferences and attitudes towards the treatment.

Literature background

Direct corrective feedback and metalinguistic feedback

We set out to investigate the effects of two forms of teacher-directed feedback techniques: direct corrective feedback (DCF), which entails giving learners the correct target language form; and metalinguistic corrective feedback (ME), which elicits learners’ metalinguistic knowledge by providing codes or as is the case in this study, rule reminders (MR). Regarding the advantages of DCF and ME, research evidence suggests that DCF is particularly useful for low level learners (Shintani & Ellis, 2013; van Beuningen et al., 2012) because its immediacy allows them to test previous hypotheses by comparing the desired linguistic form and the erroneous output, and its explicitness makes feedback more comprehensible (Chandler, 2003; Bitchener & Knoch, 2008; Ferris, 2009; Ferris et al., 2013). Furthermore, researchers point out that ME contributes to making incorrect target language use salient (Sampson, 2012), facilitates learning (Bitchener, 2012), and may be more valuable for EFL learners with previous grammar instruction (Ferris et al., 2013).

Notwithstanding, to the best of our knowledge, only two studies (Shintani & Ellis, 2013; Shintani et al., 2014) have investigated the differential effects of DCF and ME as single feedback variables (see Bitchener & Knoch, 2010a, 2010b for grouping feedback variables), but their findings reported contrasting effects for
DCF and ME on learners’ grammatical accuracy. On the one hand, in their study with 49 ESL low-intermediate students in a US university, Shintani and Ellis (2013) found that unlike DC learners, who showed no improvement over time, ME led to increased accuracy of the English indefinite article in an immediate new piece of writing but not in the text two weeks after feedback provision. Also, ME learners showed more understanding of the rule from the corrections and were able to use it when revising. The authors concluded that ME contributes to developing awareness of a rule or L2 explicit knowledge as opposed to revising after DCF. On the other hand, Shintani et al. (2014) looked into the effects of DCF and ME on the accuracy of use of indefinite articles and the hypothetical conditional of 140 general English learners, who were considered to be pre-intermediate. The researchers explored whether or not revision enhanced the feedback effect. Their results showed that there was a significant accuracy increase for the hypothetical conditional but not for articles. The authors concluded that although both DCF and ME were effective, the effects of DCF lasted longer. Similarly, the results indicated that the most effective group was DCF with opportunities for revision. The authors explained the contrasting results from Shintani and Ellis (2013) by pointing out that it targeted one grammatical feature, which may have been less burdensome for learners than concentrating on two linguistic forms that differ in complexity as in Shintani et al. (2014). Clearly, as significant as these findings are to broaden our current understanding of DCF and ME, more work is needed in a different instructional background with a broader focus (i.e., more targeted linguistic features).

The effects of comprehensive corrective feedback

Two main reasons can be discerned to explore the effectiveness of comprehensive CF (i.e., unfocused feedback) to contribute to text revision and L2 development. The first pedagogically-driven reason is that comprehensive CF is still a quite common error correction practice in the writing class (Lee, 2013). However, the increasing evidence that proves that providing CF can result in L2 development derives mainly from tightly controlled SLA studies that examined focused (i.e., selective) CF (e.g., Bitchener & Knoch, 2009, 2010b; Shintani et al., 2014). An important consideration, nonetheless, is that in L2 writing contexts where students learn to write in the target language (i.e., the learning-to-write dimension) and use writing as a means to learn the target language (i.e., the writing-to-learn-language dimension) (Manchón, 2011), selective CF as in the aforementioned studies may be pedagogically impractical if all linguistic inaccuracies must be remedied at the same time. The bottom line is that relevance of SLA findings for L2 writing teachers hinges on how applicable they are to the reality of the writing class (Ferris,
Thus, research that investigates L2 classroom feedback practices becomes paramount. Within the aforementioned instructional context, we stress the need for studies that address an L2 Writing-SLA interface (see Ortega, 2012; also Ferris, 2010 for a description of blended design studies), which looks into the value of written CF both as an instructional intervention to help learners successfully edit their texts (i.e., the L2 writing perspective) and as a learning tool to promote long-term L2 improvement (i.e., the SLA perspective).

The second theoretically-oriented reason is that of the studies that have investigated the effects of comprehensive CF (Ellis, Sheen, Murakami, & Takashima, 2008; Sheen, Wright, & Moldawa, 2009; Truscott & Hsu, 2008; van Beuningen et al., 2008, 2012), none has been conducted within a learning-to-write and writing-to-learn-language EFL writing class (see Kang & Han, 2015; Liu & Brown, 2015 for a meta-analysis of WCF). On the one hand, two of these studies compared the differential effect of comprehensive and selective CF with ESL learners (Sheen et al., 2009) and with EFL aviation, technology, and industrial design students (Ellis et al., 2008), but their self-acknowledged limitations (i.e., the unequal amount of feedback on articles and the unsystematic way of treating errors, respectively) imply that their conclusions about comprehensive and selective CF being equally effective (Ellis et al., 2008) or selective CF being superior to comprehensive CF (Sheen et al., 2009) should be read cautiously. Other studies investigated the effectiveness of comprehensive feedback and reached differing conclusions. For instance, in their sample of, inter alia, computer science and electrical engineering participants, Truscott and Hsu (2008) showed that comprehensive CF improved the accuracy of 47 EFL students during text revision but not in new writings. Conversely, in their tightly controlled studies, van Beuningen et al. (2008, 2012) provided compelling evidence that comprehensive CF brings about short-term improvement (i.e., from an initial writing task to its revision) and long-term improvement (i.e., from an initial task to a new writing task) in the accuracy of Dutch secondary school learners, yet their studies took place in a non-L2 writing environment. As a result, conclusive answers that are representative of the L2 writing class about the (in)effectiveness of comprehensive CF remain hard to obtain, which calls for more efforts to investigate to what extent such amount of feedback places a heavy cognitive load on writers’ ability to correct errors, which could in turn overwhelm them (Bitchener, 2008; Bitchener & Ferris, 2012; Sheen, 2007; Sheen et al., 2009). In doing so, whether comprehensive CF is overburdening or not needs to be examined by eliciting learners’ attitudes after receiving this type of feedback. As Storch and Wigglesworth (2010a) stated, “an investigation of linguistic factors alone is not enough” (p. 329). Therefore, further exploring learner factors could yield more conclusive evidence beyond the interpretation of the end product.
Learner factors

Evans, Hartshorn, McCollum, and Wolfersberger (2010) define learner factors as “everything the student brings to the learning experience” (p. 448). Among them, in this study we examine learners’ proficiency level, learners’ attitudes towards feedback, and learners’ feedback preferences. On the importance of the aforementioned variables in L2 learning, researchers acknowledge, among other things, (1) that the effectiveness of various types of error correction also needs to be determined considering learners’ proficiency level (Bitchener & Storch, 2016), (2) that students’ reactions to teacher feedback are likely to reflect their expectations and behavior in writing courses (Enginarlar, 1993), and (3) that a mismatch between what learners expect and what actually happens in the classroom may block learning (Nunan, 1986). From studies that have investigated either one of the aforementioned learner factors, insightful evidence about numerous aspects can be obtained. First, from the large research base on learners’ feedback preferences, evidence indicates that learners resent that correction is withheld (Schulz, 1996), want error-free writing (Amrhein & Nassaji, 2010; Leki, 1991), and believe that the teacher is responsible for correcting errors (Hamouda, 2011; Lee, 2005). Second, research on learners’ perceptions of written CF has found that learners construe feedback as useful based on their educational level and literacy level (Hedgcock & Lefkowitz, 1996), utilize strategies to handle CF depending on teachers’ error correction practice (Saito, 1994), and may view text revision as a beneficial task or as a punishment (Radecki & Swales, 1998). Also, researchers have demonstrated that the quality of feedback noticing depends on learners’ proficiency level (Qi & Lapkin, 2001) and that feedback processing can be affected by depth of processing and learners’ attitudes towards the feedback (Swain & Lapkin, 2002) and learners’ previous experiences (Hyland, 1988).

Notwithstanding, some areas still remain under-explored. For instance, the role of proficiency level in feedback outcomes, learners’ feedback preferences, or attitudes has received little attention. One study that did explore the latter was Lee (2008), who gave her 22 low proficiency and 36 high proficiency participants a questionnaire at the end of the school year to elicit their general reactions to the feedback they had previously received in four compositions. Her results, based on descriptive statistics, showed that learners’ attitudes towards the amount of correction and feedback types varied according to proficiency level. Lee (2008) concludes that “whether students find teacher feedback useful and effective hinges on a multitude of factors, including students’ proficiency and motivation and quality of teacher feedback, such as legibility” (p. 157). However, given that each learning situation is unique (Enginarlar, 1993), Lee’s (2008) study with secondary school Hong Kong learners is one insightful but limited study, and more studies
are needed to gain a deeper understanding of the interaction between proficiency, learner preferences, uptake from feedback, and grammatical accuracy. Also, this evidence needs to emerge from feedback studies. In this respect, just as Storch and Wigglesworth (2010a), Ferris and Roberts (2001) believe that “students’ attitudes and preferences have unfortunately been a neglected piece of information in many previous error correction studies and reviews” (Ferris & Roberts, 2001, p. 166). Therefore, current literature on written CF could benefit from a study that includes learners’ proficiency level as an interactional variable in learners’ grammatical accuracy, attitudes, and preferences.

The current study

This study aimed to generate theoretical and practical knowledge for SLA researchers and L2 practitioners pertaining to three research gaps: (a) the need for more evidence about the differential effects of comprehensive DCF and MR in a classroom-based context, (b) the lack of feedback studies that include interactional variables such as proficiency level to interpret the feedback effects, and (c) the need to widen current knowledge about the relation between proficiency level and learners’ feedback attitudes and preferences. More specifically, the study adds to the current body of literature on CF for two main reasons: (1) It examines the effects of CF on both learners’ editing and revision skills (i.e., the L2 writing perspective) and their L2 development (i.e., the SLA perspective); (2) It is conducted within an EFL learning-to-write and writing-to-learn setting with a sample that comprises English and English Teaching major students, leading to more ecologically-valid pedagogical implications.

The following research questions (RQ) guided this study:

RQ1. To what extent does grammatical accuracy during text revision and in new writings vary between low and high proficiency learners after comprehensive feedback with direct error correction and metalinguistic reminders?

RQ2. To what extent do attitudes towards comprehensive feedback with direct error correction and metalinguistic reminders vary between low and high proficiency learners?

RQ3. To what extent do feedback preferences vary between low and high proficiency learners?
Method

Participants and instructional context

A total of 91 students participated in this study, which took place in a urban public university in San José, Costa Rica. The participants comprised two different learner groups. The first group were 52 first-year students (40 female and 12 male, mean age = 24, SD = 3.76) enrolled in an English course that teaches all skills, including basic English writing. The students in this course met four days a week, three hours a day. The second group were 39 fourth-year students (27 female and 12 male, mean age = 29, SD = 8.84) enrolled in an advanced composition course where students learn how to write a research paper. Students in this course met three hours a day once a week. Out of the total number of participants, 71 were majoring in English and 20 were majoring in English Teaching. All participants were Spanish L1 EFL learners within a learning-to-write and writing-to learn language context (Manchón, 2011). Thus, as English and English Teaching majors, they were both learning English composition and using writing as a means to learn English. First-year students were randomly assigned into three groups: direct corrective feedback (DCF, n = 18), metalinguistic corrective feedback with rule reminders (MR, n = 17) and self-correction (SC, n = 17). In the same way, fourth-year students were assigned to DCF (n = 14), MR (n = 12), and SC (n = 13).

Target linguistic features

In the instructional setting where this study took place, L2 composition students not only learn how to write in English but also learn English through writing. As a result and considering the relatively short duration of each semester (i.e., 15 to 17 weeks), L2 teachers’ feedback targets all linguistic inaccuracies. For this study then, we kept this contextual characteristic in mind as well as Evans, James Hartshorn, & Strong-Krause’s (2011) claim that expecting corrections of unknown features is “wishful thinking” (p. 450). That is, learners in this study received comprehensive CF that dealt with all the linguistic categories (and subcategories) that they had been formally introduced to in previous courses. They were the following: subject-verb agreement, verb tenses (i.e., simple present tense, simple past tense, present perfect tense, present continuous), pronouns (i.e., subject pronoun, object pronoun, possessive adjective, possessive pronoun, demonstrative pronoun), modal auxiliaries (i.e., can, could, might, may, should, must, should, ought to), prepositions, fragments, word form, subject deletion, article errors (i.e., the, a), and subject repetition (i.e., unnecessary subject insertion). The treatment was comprehensive/unfocused because the amount of correction accrued with each
individual linguistic subcategory, adding up to more than 10 error types (see Liu and Brown’s (2015) recent classification of feedback studies).

Design and procedures

A week before the semester started, the first author met with the instructors to inform them about the investigation and discuss the logistics for each work session. Also, the instructors were asked to withhold any type of form-focused instruction or feedback on the targeted linguistic features for as long as the study lasted. Figure 1 illustrates the eight-week data collection process during which different activities were implemented throughout six sessions.

First, in week 1, the participants completed the profile sheet, took a proficiency test, and wrote their first writing piece (pretest). In week 2, the students had 20 minutes to do writing task 2. A week later, all DCF and MR groups received a copy of the original composition with the corresponding feedback and were asked to take ten minutes to study the corrections. The operationalization of DCF took the form of the provision of the correct word above the erroneous targeted form. Different from Shintani and Ellis (2013) and Shintani et al. (2014), MR was operationalized as the provision of rule reminders at the end of the students’ text with a number back to the line where the error occurred. After the time for studying the feedback was over, the researcher took away the copies with feedback and gave the students the uncorrected original texts. Then, they were handed out an extra sheet and asked to revise the composition taking into consideration the feedback they had just studied and using the uncorrected text as a guide. As we wanted to eliminate additional differences between conditions, the procedure was the same for all of them. Thus, in the case of learners who received no corrections, their copies were still taken away because they jotted down their own corrections while studying their text. Therefore, allowing them to keep the copy during revision would have added an undesired variable to the study: learners revising a text while looking at (i.e., copying) self-provided corrections (in the control group) and learners revising a text after having viewed other-provided corrections (in the experimental conditions). We also highlight that within the proficiency group students were randomly assigned to conditions but stayed in the same classroom. The time allotted for the text revision in all groups was 20 minutes.

The feedback conditions did not change throughout the 8-week study. In weeks 4 and 5, the participants followed the same procedures as far as text revision is concerned: they completed writing task 3 and revised their texts a week later based on the group they were assigned. In week 5, in addition to the last feedback session, the students also completed a questionnaire of attitudinal engagement and preferences immediately after they handed in their revised texts. Also, that same week four
Comprehensive corrective feedback on low and high proficiency writers

99

participants (two per learner group and one per experimental condition) participated in an interview to elicit their reactions towards the treatment. Finally, three weeks later, all learners had 20 minutes to write the last composition (delayed posttest).

Materials

Learner profile sheet
To gather the metadata, the participants completed a profile sheet that elicited aspects such as their personal information, schooling background, and English language background.

Placement test
Because being enrolled in the same classroom does not mean learners have the same proficiency level (Guénette, 2007), besides the participants’ self-reported English language proficiency, Oxford’s Quick Placement Test was administered to ascertain the learners’ proficiency level in both first- and fourth-year students. The results indicated that the mean proficiency level was elementary ($SD = .364$) (85% elementary, 15% lower intermediate) for first-year and advanced ($SD = .428$) (82% advanced, 10% very advanced, 8% upper intermediate) fourth-year students. Thus, first-year learners will hereafter be referred to as low proficiency learners (LP) and fourth-year ones as high proficiency learners (HP).

<table>
<thead>
<tr>
<th>Week</th>
<th>Session</th>
<th>Time</th>
<th>Direct CF (DCF)</th>
<th>Metalinguistic CF with rule reminders (MR)</th>
<th>Self-correction (SC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>5</td>
<td>Learner profile sheet</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>Writing task 1: pretest (“Social Networks”)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>20</td>
<td>Writing task 2 (“Education and Student Life”)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>10</td>
<td>Studying text based on DCF</td>
<td>Studying text based on MR</td>
<td>Studying text based on SC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>Revising text</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>20</td>
<td>Writing task 3 (“Money”)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>10</td>
<td>Studying text based on DCF</td>
<td>Studying text based on MR</td>
<td>Studying text based on SC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>Revising text</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>Completing questionnaire</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>Interview</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>20</td>
<td>Writing task 4: delayed posttest (“Dream Jobs”)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Study design. DCF = direct CF, MR = metalinguistic CF with rule reminders, SC = self-correction with no CF
Writing tasks
Since the School of Modern Languages where this investigation took place has scarce computer lab facilities for composition courses, all texts were pen-and-paper in-class compositions. Also, as the study was implemented in actual classes and in existing courses, both class time and the pedagogical demands of the courses were decisive for determining the time on task, the writing prompts, and the topics – hence the ecological validity of this study. First, because part of the writing course goals was expressing one’s opinion, the nature of the task was argumentative writing. In this sense, the writing tasks were designed in a way in which learners had some freedom of expression as Bruton (2009b) suggests. Second, in line with the course materials, the topics in all sessions were chapter-based. Therefore, students wrote about Social Networks, Education and Student Life, Money, and Dream Jobs. Third, on the basis of class time and prior experience of the teachers and teacher/researcher with the specific target audience, the time allotted for the compositions was 20 minutes, which was considered a feasible time span to complete the writing tasks that were required. In general, for all topics, students received a short reading excerpt with two prompt questions below. Question 1 was the same in all writing tasks and question 2 followed up with a reading-based prompt (see Appendix A). The topics could not be counterbalanced because they were based on the course syllabus. However, as students were familiarized with these topics during the course, we did not expect topic knowledge to play a decisive role. Consequently, the tasks were considered to be similar in level of complexity. Additionally, we also took into account possible tasks effects no matter how unlikely in our statistical analyses.

Questionnaire
We administered a questionnaire of attitudinal engagement and feedback preferences given the need for more empirical evidence emerging from feedback studies on affective variables such as feedback preferences and attitudes towards specific feedback strategies (Ferris & Roberts, 2001; Storch & Wigglesworth, 2010a). Therefore, a total of 73 students anonymously completed the instrument ($\alpha = .84$), which was administered immediately after the feedback to shorten the reference period and avoid biased feedback estimates (De Leeuw & Dillman, 2008). The questionnaire comprised three parts (see Appendix B):

A. a 4-point Likert scale to elicit learners’ response to the treatment,
B. a list of eight CF techniques to probe learners’ preferred feedback type, and
C. two sets of items to elicit learners’ feedback preferences concerning degree of explicitness and amount of feedback.
In order to design the items for Part A, we drew, on the one hand, on studies by Leki (2006) and Enginarlar (1993), and on an extensive review of the literature and previous claims against comprehensive CF, on the other hand (e.g., Sheen et al. 2009). Since a large number of studies probe learners’ general perspectives on CF (e.g., Incecay & Dollar, 2011; Loewen et al., 2009; Schulz, 1996) and not specific reactions to a particular treatment that had been just provided, we were not able to base ourselves on existing, validated instruments. Hence, Part A had 11 items that measured learners’ reactions to the feedback (see Coding and analysis for factor loadings). Next to each statement, learners had to indicate their level of agreement from 1 (strongly agree) to 4 (strongly disagree). Below each statement they also had the option to explain the reason(s) for their answers.

Then, for eliciting learners’ preferences (Parts B and C), we drew on the questionnaire by Leki (1991) and Lee (2008). More specifically, for Part B the choice of the feedback techniques was based on Ellis (2009) typology of feedback types. Finally, influenced by the context of this investigation (i.e., actual L2 writing classes and in existing courses), we decided to elicit both learners’ attitudes towards the feedback and learners’ feedback preferences in a single data collection tool. In other words, for the design we also had to make sure that the response time of the questionnaire could fit the class time and schedule restrictions.

Interview
To add further depth to the study concerning learners’ attitudes towards the feedback, we conducted an interview with four students (see interview protocol in Appendix C). The four learners (two LP and two HP students, distributed over the two conditions) agreed to participate in a semi-structured interview on a voluntary basis (adapted from Leki, 2006). Also, they gave consent to record the interview, which took place after the last feedback session. Afterwards, the first author transcribed the audio.

Coding and analysis
Once the information from the learner profile sheet and the proficiency tests was obtained, all compositions were tagged to ensure anonymity of the texts. All hand-written texts were transcribed using Dragon Naturally Speaking 11.0., which was needed for both avoiding assessment bias and obtaining word counts. In cases in which the aforementioned program could not transcribe an error, we inserted it manually. Once the word-processing text conversion was finished, the texts were coded for the targeted linguistic features. After tracing the errors (see Bruton, 2009; Storch, 2010 for assessment of accuracy changes) and keeping in mind that absence of error does not necessarily mean successful error correction (Murphy & Roca de Larios, 2011), four variables for revision behavior were computed per
revision session: the number of errors that were successfully revised divided by the total number of errors (i.e., proportion corrected), the number of errors that were kept divided by the total number of errors (i.e., proportion maintained), the number of errors that were incorrectly revised divided by the total number of errors (i.e., proportion wrongly substituted), and the number of errors that were deleted from the text divided by the total number of errors (i.e., proportion avoided).

To determine the reliability of the error counts, three raters recoded the same 40 texts from the post-test session, which were randomly selected from all the groups. Cronbach’s Alpha for the raters on the global error count and ten linguistic features was .958 and .876, respectively. As for the overall accuracy measures of the pretest and posttest, as did van Beuningen et al. (2008, 2012), we used an error-words ratio (i.e., \([\text{number of grammatical errors/total number of words}] \times 10\)).

Regarding the statistical analyses, once the assumptions of homogeneity and normality were confirmed, we conducted a two-way ANOVA for each session separately to test for condition differences for the two proficiency groups for the two revision sessions. The full model included condition and level as between factors, and the interaction between condition and level. Furthermore, to determine whether a possible long-term effect was due to the experimental condition, we used a two-way ANCOVA on the overall posttest accuracy measure with the overall pre-test accuracy measure as a covariate. AN(C)OVA was considered to be the most suitable statistical test for the current study considering its (imbalanced) task design and the fact that even though the different prompts/tests are considered as similar (and were vetted by teachers and experts alike), their level of difficulty is unknown, which means that “effects of measurement occasion cannot be interpreted as developmental” (Groenendijk et al., 2013, p. 41) and that we can only examine the effects on the different tests separately (cf. Bouwer, Béguin, Sanders, & van den Bergh, 2014; Groenendijk et al., 2013). Effect sizes were calculated as eta-squared (\(\eta^2_p\)) with values of \(0.01 < \eta^2_p < 0.06\) (small), \(0.06 < \eta^2_p < 0.14\) (moderate), and \(0.14 < \eta^2_p\) (large) (Cohen, 1988). Concerning the questionnaire, we subjected the quantitative items to an exploratory factor analysis with a Varimax (orthogonal) rotation, which yielded a determinant value of .005, a Kaiser-Meyer Olkin (KMO) measure of .798, and a highly significant Bartlett’s test (\(p = .000\)). We opted for an exploratory factor analysis given the lack of theory emerging from feedback studies about learners’ attitudes towards teacher feedback – as acknowledged in Ferris and Roberts (2001), Hyland and Hyland (2006), and Storch (2010). The exploratory factor analysis indicated that the scale items of Part A loaded on three components that dealt with the following: learners’ general perception of the overall benefits of the feedback, learners’ feelings regarding the suitability of the feedback, and learners’ self-perceived grammatical improvement as a result of the feedback. Consequently, we labeled the constructs feedback utility, feedback
appropriateness, and self-perceived grammatical improvement, respectively (see Table 1). In a similar way, the qualitative data gathered from learners’ written comments in the questionnaire were categorized in accordance with the constructs and contrasted by level.

Table 1: Summary of components based on factor loadings for exploratory factor analysis

<table>
<thead>
<tr>
<th>Feedback utility</th>
<th>Feedback appropriateness</th>
<th>Self-perceived grammatical improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 3 (.858)</td>
<td>Item 6 (.754)</td>
<td>Item 8 (.898)</td>
</tr>
<tr>
<td>Item 7 (.767)</td>
<td>Item 5 (.698)</td>
<td>Item 4 (.834)</td>
</tr>
<tr>
<td>Item 2 (.711)</td>
<td>Item 10 (.638)</td>
<td></td>
</tr>
<tr>
<td>Item 9 (.663)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1 (.631)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 11 (.488)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As for the qualitative data from the interviews, they were analyzed following Creswell’s (2008) model of qualitative coding process. After that, because one strategy to achieve qualitative validity is having others examine the data independently (Creswell and Clark, 2011), two experienced writing researchers reviewed the data. The double check by two other researchers was meant to test and validate our coding system. Hence, on two occasions the procedure involved reflection and deep discussion, which led to a more precise and fine-grained coding. Consequently, after careful analysis, the resulting extracted themes from learners’ scale answers were the following: feedback quality, feedback benefits, feedback appropriateness, learners’ wants, writing self-rating, error detection ability, and views on written CF.

As for the statistical tests, the variables of the answers for Part A of the questionnaire were measured on an ordinal scale. Thus, we used a one-way ANOVA to determine to what extent learners’ proficiency level and attitudes towards the treatment differed. For parts B and C of the questionnaire, we ran a chi-square test of independence comparing the feedback preferences in LP and HP since the variables were measured on a nominal scale. Hence, in case we obtained an unexpected count of at least one cell lower than 5, we interpreted the exact test otherwise Asymptotic version of this test (Field, 2013).
Results

The research questions explored the effects of comprehensive DCF and MR on immediate accuracy and accuracy improvement (RQ1) and learners’ attitudes towards the feedback (RQ2). They also looked into learners’ feedback preferences (RQ3). The results for the three research questions will be reported successively in this section.

Immediate accuracy

Table 2 shows the descriptive statistics for revision behavior by condition, level, and session. Inferential statistics on these data yielded the following results. A two-way ANOVA showed general statistically significant effects for condition in the number of errors corrected during revision of two writing tasks: writing

Table 2. Summary of descriptive statistics for revision behavior by condition, level and session

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Group</th>
<th>LP WT2</th>
<th>LP WT3</th>
<th>HP WT2</th>
<th>HP WT3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected</td>
<td>DCF</td>
<td>17 .823 .139</td>
<td>17 .830 .243</td>
<td>11 .746 .418</td>
<td>12 .041 .097</td>
</tr>
<tr>
<td>Total</td>
<td>48 .618 .360</td>
<td>48 .606 .344</td>
<td>34 .693 .422</td>
<td>28 .089 .227</td>
<td></td>
</tr>
<tr>
<td>Maintained</td>
<td>DCF</td>
<td>17 .156 .139</td>
<td>17 .080 .118</td>
<td>11 .065 .192</td>
<td>12 .625 .410</td>
</tr>
<tr>
<td>SC</td>
<td>15 .749 .260</td>
<td>15 .552 .384</td>
<td>11 .1.25 2.61</td>
<td>7 .285 .393</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>48 .332 .341</td>
<td>48 .265 .315</td>
<td>34 .422 1.52</td>
<td>28 .577 .460</td>
<td></td>
</tr>
<tr>
<td>Substituted</td>
<td>DCF</td>
<td>17 .000 .000</td>
<td>17 .014 .060</td>
<td>11 .000 .000</td>
<td>12 .000 .000</td>
</tr>
<tr>
<td>SC</td>
<td>15 .000 .000</td>
<td>15 .016 .045</td>
<td>11 .000 .000</td>
<td>7 .000 .000</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>48 .002 .016</td>
<td>48 .016 .051</td>
<td>34 .004 .028</td>
<td>28 .011 .062</td>
<td></td>
</tr>
<tr>
<td>Avoided</td>
<td>DCF</td>
<td>17 .019 .080</td>
<td>17 .075 .188</td>
<td>11 .000 .000</td>
<td>12 .000 .000</td>
</tr>
<tr>
<td>MR</td>
<td>16 .015 .062</td>
<td>16 .104 .188</td>
<td>12 .020 .072</td>
<td>9 .000 .000</td>
<td></td>
</tr>
<tr>
<td>SC</td>
<td>15 .131 .269</td>
<td>15 .163 .279</td>
<td>11 .000 .000</td>
<td>7 .000 .000</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>48 .053 .167</td>
<td>48 .112 .219</td>
<td>34 .007 .042</td>
<td>28 .000 .000</td>
<td></td>
</tr>
</tbody>
</table>

Note. LP = low proficiency learners; HP = high proficiency learners; WT2 = writing task 2; WT3 = writing task 3; DCF = direct corrective feedback; MR = metalinguistic reminders; SC = self-correction.
task 2 (WT2), \([F(2, 82) = 25,107, p = .000, \eta_p^2 = .39]\); and writing task 3 (WT3), \([F(2, 76) = 12,440, p = .000, \eta_p^2 = .26]\).

Post hoc comparisons using Bonferroni showed that the number of errors corrected did not differ significantly between DCF and ME during revision of WT2 \((p = 1.00)\) and WT3 \((p = 1.00)\). However, both treatments did significantly outperform the control group (SC) during revision of WT2 \((p = .000)\) and WT3 \((p = .000)\) with very large effects \((\eta_p^2 = .39\) and \(\eta_p^2 = .26\), respectively\). This means DCF and MR were equally effective in helping learners to produce a correct version of their errors. Table 3 provides an overview of the significant contrasts for immediate accuracy with associated Cohen’s \(d\) effect sizes per writing task and level. As can be seen, LP and HP learners who received DCF and MR had a larger accuracy advantage during both revision tasks (i.e., WT2 and WT3) over students in the SC groups.

Table 3. Summary of significant accuracy contrasts for immediate accuracy with associated Cohen’s \(d\) effect sizes

<table>
<thead>
<tr>
<th></th>
<th>LP</th>
<th>HP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate accuracy Corrected in WT2</td>
<td>DCF&gt;SC 3.55</td>
<td>DCF&gt;SC 0.65</td>
</tr>
<tr>
<td></td>
<td>MR&gt;SC 2.98</td>
<td>MR&gt;SC 1.02</td>
</tr>
<tr>
<td></td>
<td>DCF&gt;SC 2.11</td>
<td>DCF&gt;SC 0.78</td>
</tr>
<tr>
<td></td>
<td>MR&gt;SC 1.63</td>
<td>MR&gt;SC 1.10</td>
</tr>
</tbody>
</table>

Note. WT2 = writing task 2, WT3 = writing task 3, LP = low proficiency learners, HP = high proficiency learners, DCF = direct corrective feedback; MR = metalinguistic reminders; SC = self-correction.

An important aspect to highlight is that even though the difference between the experimental conditions did not reach statistical significance in WT2 and WT3 \((p = .1000)\), learners’ grammatical correctness during both revisions benefited differently from each condition in each level. The marginal means (see Figure 2) show that for LP learners, DCF (Cohen’s \(d = 3.55\)) brought about more immediate accuracy in WT2 than MR (Cohen’s \(d = 2.98\)). The same holds for WT3 where DCF (Cohen’s \(d = 2.11\)) turned out to be more beneficial than MR (Cohen’s \(d = 1.63\)). For HP learners, conversely, MR (Cohen’s \(d = 1.02\)) helped them more to successfully correct their errors in WT2 than DCF (Cohen’s \(d = 0.65\)). Similarly, in WT3 learners benefited more from ME (Cohen’s \(d = 1.10\)) than DCF (Cohen’s \(d = 0.78\)).

Furthermore, a two-way ANOVA showed general statistically significant effects for condition in maintained errors during revision of WT2, \([F(2, 82) = 7,460, p = .001, \eta_p^2 = .16]\); and maintained errors during revision of WT3, \([F(2, 76) = 15,167, p = .000, \eta_p^2 = .30]\). Bonferroni post-hoc tests indicated that the two experimental conditions significantly differed from the control group in the number of errors maintained during text revision of WT2 \((p = .005)\) and
WT3 ($p = .000$). More specifically, in WT2 LP learners that self-corrected their texts made significantly fewer corrections than those who received DCF ($p = .005$, Cohen’s $d = 2.84$) and MR ($p = .005$, Cohen’s $d = 2.49$). Also, in WT3 the same learners maintained significantly more errors than those who corrected them with DCF ($p = .000$, Cohen’s $d = 1.67$) or with MR ($p = .000$, Cohen’s $d = 1.20$). The same results were obtained for HP students in the control group, who corrected significantly less in WT2 than those in the DCF ($p = .005$, Cohen’s $d = 0.61$) and MR ($p = .005$, Cohen’s $d = 0.66$) groups. In WT3, self-correcting their texts made
HP learners maintain significantly more errors than receiving DCF ($p = .000$, Cohen’s $d = 0.85$) and MR ($p = .000$, Cohen’s $d = 1.02$).

As for substituted and avoided errors, there were no statistically significant effects of group condition for the number of wrongly substituted errors during revision of WT2, $[F(2, 82) = .2000, p = .142, \eta_p^2 = .05]$; wrongly substituted errors during revision of WT3, $[F(2, 76) = .093, p = .410, \eta_p^2 = .02]$; avoided errors during revision of WT2, $[F(2, 82) = 1.437, p = .244, \eta_p^2 = .03]$; and avoided errors during revision of WT3, $[F(2, 76) = .356, p = .702, \eta_p^2 = .01]$.

Concerning the interaction between condition and level, it did not reach statistical significance at the $p < .05$ level in any of the revision behaviors of both writing sessions: corrected errors in WT2, $[F(2, 82) = .2597, p = .081, \eta_p^2 = .064]$; and WT3, $[F(2, 76) = 1.138, p = .326, \eta_p^2 = .031]$; maintained errors in WT2, $[F(2, 82) = .867, p = .424, \eta_p^2 = .022]$; and WT3 $[F(2, 76) = 1.502, p = .230, \eta_p^2 = .041]$; wrongly substituted errors in WT2, $[F(2, 82) = .222, p = .801, \eta_p^2 = .006]$; and WT3, $[F(2, 76) = .773, p = .465, \eta_p^2 = .022]$; and avoided errors in WT2, $[F(2, 82) = 2.107, p = .129, \eta_p^2 = .053]$; and WT3, $[F(2, 76) = .356, p = .702, \eta_p^2 = .010]$. This indicates that in this study the condition effect on revision behavior does not depend on proficiency level.

**Accuracy improvement**

Regarding the long-term effects of the treatment, after a two-way ANOVA yielded a main effect for condition in the overall post-test accuracy measure $[F(2, 75) = 3.558, p = .034, \eta_p^2 = .09]$, we conducted a two-way ANCOVA on the overall post-test accuracy measure with the overall pre-test accuracy measure as a covariate in order to determine the effect of the feedback to improve grammatical accuracy in new writings.

The test showed that there was a significant effect for condition $[F(2, 75) = 5.083, p = .009, \eta_p^2 = .13]$. Planned contrasts (Field, 2013) indicated that receiving comprehensive DCF ($p = .002$, 95% CI [$-.178, -.040$]) and MR ($p = .036$, 95% CI [$-.148, -.005$]), significantly contributed to long-term accuracy improvement compared to not receiving any correction at all. LP learners’ grammatical accuracy in the DCF (Cohen’s $d = 0.83$) and the MR (Cohen’s $d = 0.89$) groups significantly improved more than those in SC. In like manner, HP students that received DCF (Cohen’s $d = 0.72$) and MR (Cohen’s $d = 0.20$) significantly outperformed the control group. We did not find a statistically significant effect for level $[F(1, 75) = .602, p = .440, \eta_p^2 = .009]$ or the interaction between level and condition, $[F(2, 75) = 2.373, p = .101, \eta_p^2 = .065]$.

Overall, from these results and those from the previous Section five aspects are worth noting: (1) both DCF and MR were effective to increase LP and HP learners’
grammatical accuracy during text revision and in new writings since both outperformed the SC group, (2) the general effect size of comprehensive DCF and MR was large during text revision and moderate in new writings, (3) the magnitude of the effect of either feedback type proved to be greater on LP learners than on their HP counterparts during text revision and in new writings, (4) the condition effect on learners’ grammatical accuracy during text revision and in new writings did not depend on proficiency level, and (5) the difference between DCF and MR did not reach statistical significance in two revision tasks, yet an emerging pattern showed that during text revision, LP students achieved more successful error correction with DCF while HP reached more grammatical accuracy with MR.

Learners’ attitudes

Quantitative results

Table 4 shows the descriptive statistics for attitudes by level and construct. A one-way ANOVA was conducted to determine to what extent low and high proficiency learners’ attitudes towards the feedback differed.

Table 4. Summary of descriptive statistics for attitudes by level and construct

<table>
<thead>
<tr>
<th>Level</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Self-perceived grammatical improvement</td>
<td>Feedback appropriateness</td>
<td>Feedback utility</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LP</td>
<td>48</td>
<td>3.81</td>
<td>1.21</td>
<td>5.75</td>
<td>1.39</td>
<td>9.75</td>
<td>2.69</td>
</tr>
<tr>
<td>HP</td>
<td>25</td>
<td>4.20</td>
<td>1.38</td>
<td>6.44</td>
<td>1.00</td>
<td>9.52</td>
<td>2.78</td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
<td>3.94</td>
<td>1.27</td>
<td>5.98</td>
<td>1.30</td>
<td>9.67</td>
<td>2.70</td>
</tr>
</tbody>
</table>

Note. LP = low proficiency learners, HP=high proficiency learners.

The results of this test for the three constructs indicated that there was a statistically significant level effect for feedback appropriateness, $F(1, 73) = 4.825, p = .031, \eta^2_p = .064$, which implies there is a relation between learners’ proficiency level and their attitudes towards the feedback. However, no statistical significance was found for feedback utility, $F(1, 73) = .117, p = .733, \eta^2_p = .002$; or self-perceived grammatical improvement, $F(1, 73) = 1.520, p = .222, \eta^2_p = .021$. Overall, a closer look at learners’ reactions to the feedback reveals that LP and HP students were most in agreement with their perception of feedback utility (e.g., how the feedback was not only clear but also helpful to understand what was wrong) and grammatical improvement (e.g., the majority found it necessary to improve their accuracy).
Qualitative results
The positive appreciation towards feedback utility and the perceived grammatical improvement was further supported in the answers from the questionnaire where learners could elaborate on their attitudes if desired. In justifying their reactions towards the aforementioned constructs, both LP and HP students recurrently wrote reasons such as the following:

1. It allowed me to see my written errors and correct them.
2. It helped me to improve.
3. It helped me to notice my mistakes.

Interestingly, although in both learner groups there was a general welcoming attitude pertaining to the utility of the feedback and its aid to improve grammatical accuracy, some LP and HP learners agreed they wished they had been corrected differently. This desire for something different – despite acknowledgement of feedback usefulness – was expressed in LP students’ comments such as:

4. But I’m starting my English studies, so any feedback is good feedback.
5. I wanted more in other areas like content.
6. Maybe I could use an oral explanation from the teacher.

Along the same lines, HP learners stated:

7. It was not challenging but it was helpful.
8. Any feedback is helpful.
9. Maybe it could focus not only on grammar but on content.

Regarding feedback appropriateness, the underlying reasons behind learners’ attitudes reveal that their self-perceived proficiency level may have played a role in the discrepancy between the two learner groups. For instance, on the reasons why the feedback was not appropriate, HP students wrote:

10. The feedback gave me the correct way to do so, but I must be able to find the answer on my own.
11. We have taken all grammar courses.
12. We’re in fourth year.
13. I believe my grammar is at a good level.

Conversely, LP learners expressed the feedback was appropriate because:

15. It’s right for what we already know.
16. We can’t self-correct all errors.
Furthermore, we could obtain more insight into learners’ attitudes towards the feedback from the interview answers (see summarized answers in Appendix D). Just as in the questionnaire comments, learners’ self-perceived proficiency level emerged in the interviews. For example, the HP student who received MR affirmed that direct corrections are “not the ideal because they are giving the answer and … the idea is for you to analyze which is the correct answer”. Still, she acknowledged that in her first year of the major, it was “good because I was just starting with the composition courses”. Nevertheless, when asked about the possibility of receiving such feedback type as a fourth-year student, she replied that she “would feel, not offended, but I would feel like the professor doesn’t believe I’m capable of making the correction by myself, so no, I wouldn’t like that”. For this reason, despite the explicitness of MR, she still found it helpful because she did not get “the actual answer”.

Similarly, the LP student who felt that DCF did not “help that much” went on to say that she still valued it because “my English level isn’t so advanced to allow me to say that no feedback helps me at this point”. In fact, her perspective of the feedback benefits can be understood in her elaboration of her feedback needs: she wants more emphasis on content and organization, so feedback on grammatical issues is not enough. Conversely, the remaining learners appreciated the quality and benefits of the feedback more although they needed feedback on other issues. When elaborating on their views on written CF, all learners acknowledged its importance, but their elaboration in three out of four cases was followed by “but”, which introduced their feedback preferences. Thus, for the HP student that received DCF, whether or not written CF helps “depends on the kind of feedback” because “when they give me a lot of codes, it’s kind of distracting to me.” Therefore, he prefers “to have a circle and then I can figure out what’s wrong”. Instead, for the HP learner who corrected her errors with MR, written CF is helpful “in some way, but after talking to the professor” because “there has to be a balance between the written feedback and the professor.” In the case of the LP student, the one who received MR added that “it [giving feedback] should be done with oral explanation”.

Learners’ feedback preferences

A chi-square test of independence examined the association between proficiency level and preferred feedback techniques (see summary in Appendix E). The results revealed a statistically significant relation for metalinguistic feedback with codes, $\chi^2 (1, = 69) = 5.82, p = .016$. This means that HP learners are more likely to favor this type of feedback than LP learners. However, marginal feedback [$\chi^2 (1, = 69) = 1.786, p = .181$], indirect feedback with underlining (or circling) [$\chi^2 (1, = 69) = .233, p = .630$], metalinguistic feedback with rule reminders [$\chi^2$]
Comprehensive corrective feedback on low and high proficiency writers

(1, = 69) = .759, p = .384], direct CF [χ² (1, = 69) = .053, p = .818], direct CF plus metalinguistic codes or rules reminders [χ² (1, = 69) = .224, p = .636], electronic feedback [χ² (1, = 69) = 1.170, p = .279], and reformulation [χ² (1, = 69) = .198, p = .656] did not yield a statistically significant association. These nonsignificant results imply that both LP and HP learners equally prefer other feedback types. Still, a closer look at learners’ preferences shows that while HP students considered feedback types that require a higher depth of processing ranging from CF with codes and marginal feedback to underlining and reformulation, LP students did so to a lesser degree. Instead, the latter leaned more towards feedback strategies that rely heavily on the teacher and require a minimal depth of processing (e.g., underlining + correct form + grammar rule) (cf. Appendix E).

We also performed a chi-square test of independence on learners’ preferences concerning degree of explicitness and feedback amount (see summary in Appendix F). The test did not yield a statistically significant relation with preferred degree of feedback explicitness, [χ² (1, = 73) = 3.489, p = .062], but it did with learners’ preferred amount of feedback [χ² (1, = 73) = 6.578, p = .010]. The latter result points to the likelihood of LP learners favoring a comprehensive approach to errors more than their HP counterparts.

Discussion

Overall, the contribution of our study to the current research base on CF relates to two domains: L2 writing and SLA. In this section, we will then interpret the results concerning grammatical accuracy, learners’ attitudinal engagement with feedback, and learners’ feedback preferences and discuss them with reference to their contribution and implications.

Grammatical accuracy

Regarding the extent to which comprehensive DCF and ME during text revision can enhance correctness (RQ1), our results demonstrate that learners’ grammatical correctness was significantly greater for the experimental groups, who could successfully correct more and maintain fewer errors than those who got the opportunity to self-correct their text. These findings are similar to previous studies that included revision in their design, where learners who received CF outperformed those who did not (e.g., Ferris, 2006; Ferris & Roberts, 2001; Truscott & Hsu, 2008). Also, as did van Beuningen et al. (2008, 2012), we found evidence that comprehensive CF can facilitate accuracy in the writings of new texts (RQ1). However, our study in particular shows that feedback provided on more than one
occasion (vis-à-vis one-shot treatment) can enhance low and high proficiency learners’ grammatical accuracy during text revision and in subsequent writings. The learning potential of the feedback is then evident in learners’ sustained grammatical accuracy over time (i.e., improved long-term changes in performance), which constitutes evidence of uptake for Bitchener (2012) or acquisition for Polio (2012). Our results do not concur with those of Shintani & Ellis (2013) and Shintani et al. (2014) in that both DCF and MR were equally effective in enhancing short and long-term accuracy. We can explain these differences in terms of this study’s population and the operationalization of MR. Following the explanation put forward by Shintani et al., (2014) where they stated that their participants “had substantial metalinguistic knowledge of English grammar” (p. 8) based on their students’ years of formal English instruction in high school and at the university, we hypothesize that our results may be, in part, due to metalinguistic knowledge differences. That is, while the participants in Shintani and Ellis (2013) and Shintani et al. (2014) were ESL students and FL learners majoring in mathematics, science, music, physical education, arts, special needs education, and English, respectively, the learners in our study were majoring in English or English teaching. As a result, considering their particular instructional context and former years of English exposure in high school, it seems plausible that the participants in our study may have had a higher metalinguistic knowledge base than those in Shintani and Ellis (2013) and Shintani et al. (2014). Hence, it is likely that LP and HP learners’ ability to notice more effectively the gap between their output and the L2 input in the form of CF was enhanced by the saliency and explicitness of individualized MR and the immediacy of DCF, to the extent that they used that linguistic knowledge for accurate revision and subsequent retrieval in new writings.

Conversely, the generic operationalization of ME in Shintani and Ellis (2013) and Shintani et al. (2014), which required a different level of engagement and cognitive effort (i.e., a handout with rules vis-à-vis in-text corrections) may have prevented learners – whose metalinguistic knowledge base may have been lower – from confirming their hypotheses in a more timely manner even when one or two grammatical features were treated. This explanation fits research that indicates that the effectiveness of explicit CF types depends on learners’ metalinguistic knowledge background (Bitchener & Knoch, 2008; Ferris & Roberts, 2001). However, we acknowledge that measuring learners’ metalinguistic background knowledge is deemed desirable in future studies to further substantiate this claim. Also, we cannot overlook the fact that in Shintani et al. (2014) the two targeted features responded to correction differently. Hence, differences might have also occurred due to characteristics of the linguistic features in our study and those of Shintani et al. (2014) and Shintani and Ellis (2013), which differed not only in foci but also in degree of complexity.
As far as grammatical accuracy is concerned, from an SLA standpoint our findings prove that comprehensive CF with DCF and ME have the potential to facilitate L2 development. Cognitively speaking (see stages of cognitive processing of input by Gass, 1997 in Bitchener & Storch, 2016, p. 18), the L2 learning evidence in our study suggests that our participants were able to consciously attend the feedback, notice (with understanding) the discrepancy between the input (in the form of CF) and their output, match the input with their existing stored linguistic knowledge (i.e., intake), process it, and produce accurate, modified L2 output in new writings. However, it should be noted, as Bitchener and Storch (2016) explain, that when there is evidence of L2 development, caution must be exercised as to the claims that are made. When learners are able to improve linguistic accuracy over time, it means that they have successfully completed “the stages in the development of L2, from the initial written CF input stage to the implicit, automatized output stage” (p. 2). Nevertheless, they warn readers that more opportunities are needed to consolidate that L2 knowledge before claims about L2 acquisition – vis-à-vis L2 development – are made. That is to say, whereas the term L2 development refers to the stages learners follow to develop such knowledge, L2 acquisition refers to “the acquired end-product (native speaker mastery and competence)” (p. 2).

Second, from an L2 writing perspective, L2 teachers looking for pedagogical answers for their error correction practice need more evidence about the usefulness of comprehensive DCF and ME as editing and revision tools when implemented with low and high proficiency learners such as the participants in this study. It seems logical, though, that some L2 composition teachers could feel reluctant to implement DCF in their class since as Polio (2012) rightly points out, “It is obvious that a writer can look at direct corrections and copy them onto a new piece of paper. Even if those corrections were indirect, the second piece of writing would be more grammatically accurate” (p. 377). Nevertheless, to bring about immediate accuracy and long-term accuracy improvement, the type of DCF (and MR) that we suggest is one where learners study the feedback, have the corrections taken away from them, and revise the text with a clean copy – as operationalized in this study. Our findings demonstrate that going about feedback provision in this way is a fruitful and pedagogical error correction practice, which may contribute to reducing the probabilities of CF going unattended. In fact, revision is considered pushed output (Polio, 2012) and “likely a necessary, or at least helpful, step on the road to longer term improvement in accuracy” (Ferris, 2004, p. 54).

Indeed, we can infer from low and high learners’ sustained accuracy and their general appreciation of the treatment, it seems that comprehensive feedback as provided in our study does not produce a cognitive overload. However, from existing literature on CF, it still remains unclear how many linguistic categories are too many for learners to process effectively and to what extent that hinges on cognitive
factors. If 15 linguistic features is too many (Ferris, 2010), would learners attending all surface-level issues (i.e., grammar, spelling, punctuation, capitalization), for example, feel overwhelmed? If so, could that tax their ability to process the feedback? Can learners’ grammatical accuracy be affected when their attention is drawn to other issues? Certainly, both the LP and HP learners in our study demonstrated that they could affectively (on the basis of self-report data) and cognitively (on the basis of product data) handle feedback targeting a large array of errors, but the questions above warrant more research that explores to what extent a broader approach would render similar results concerning feedback effects, learners’ attitudinal engagement, and cognitive load due to feedback. A study examining this issue is already underway.

We did not find any evidence in this study that the condition effect on learners’ grammatical accuracy depends on proficiency level. Nevertheless, the emerging pattern in our results, where learners with low proficiency were more grammatically accurate with direct provision of the target language forms and those with high proficiency with metalinguistic reminders, is worth investigating in the future. Previously, claims have been made about direct error correction being more useful for low level learners (Ferris & Roberts, 2001; Ferris & Hedgcock, 2005) and metalinguistic feedback for EFL learners with a (more advanced) metalinguistic background (Ferris et al., 2013). However, more empirical validation is needed before firm answers can be drawn. Such research effort is welcome not only theoretically but also practically, especially considering L2 instructional contexts where all learners are corrected the same way regardless of their proficiency level.

Attitudinal engagement

We found evidence that there is a relation between proficiency level and learners’ attitudes towards the feedback (RQ2). The statistically significant difference between LP and HP students pertaining to feedback appropriateness together with comments in the questionnaire and answers from the interviews suggest that HP learners’ attitudes towards the feedback may have been influenced by their self-perceived level of grammatical competence. This hints at the possibility that even though HP students did not strongly resent the treatment as evidenced in their answers about feedback utility and self-perceived grammatical accuracy, it did give them more corrective information than they thought they needed. Conversely, LP learners indicated they wanted more (e.g., “I think I needed more”). Taken together, our results imply that attitudinal differences towards feedback can arise due to a lack of alignment between teachers’ feedback practices and learners’ self-perceived grammatical competence. In this sense, our findings are in line with previous studies that state that learners’ needs “become more and more individualized” as
learners progress (Evans et al., 2010, p. 450) and that their motivation may lessen if they do not get what they want (Ferris, 2003 in Hyland & Hyland, 2006b).

Interestingly, some LP and HP learners wished they had been corrected differently even though they showed a general positive appreciation of the feedback. This could be construed as incongruent. However, it could be explained as follows: When learners perceive some benefits in their grammatical accuracy as a result of feedback, their perception of its usefulness may not be negatively affected even if the treatment does not entirely match their needs. If this is true, acknowledgment of feedback usefulness is not indicative of preference. Indeed, that students may see value in feedback types that do not necessarily fit their real feedback preferences can be further explained by considering the context surrounding the feedback provided in this study: that of EFL where the target language learning opportunities outside the academic environment are limited (Hedgcock & Lefkowitz, 1996). Indeed, Hedgcock & Lefkowitz (1994) rightly stated, “Learners’ perceptions about what constitutes useful feedback vary according to the orientation and demands of the specific L2 learning context” (p. 157). For this very reason, the EFL learners in this study may be aware that CF is one of the limited sources of expert input surrounding them and as such, it is welcome if it aids their linguistic accuracy – especially considering that succeeding as English (Teaching) majors and English (Teaching) professionals may largely depend on their target language mastery.

Overall, we can draw significant implications from our results. First, given that an association between learners’ proficiency level and feedback attitudes (and preferences) exists – as our findings showed – and that affective factors can influence the way learners engage with feedback and benefit from it (Bitchener & Storch, 2016; Lee, 2013), L2 teachers may want to make an effort to uncover their students’ preferences (see p. 215 on Hyland & Hyland, 2006b for suggestions), which could in turn influence their attitude towards CF (Ellis, Sheen, Murakami, & Takashima, 2008). Second, to prevent (potential) negative attitudes towards the feedback, L2 teachers may also want to reflect on the importance of letting learners’ proficiency level and the instructional context that surrounds the feedback factor into their feedback decisions and practices.

**Feedback preferences**

Our results showed that an association between proficiency level and learners’ feedback preferences exists (RQ3). Our findings suggest that HP learners are more likely to favor metalinguistic feedback with codes than LP students. This difference can be explained in conjunction with the learners’ instructional setting. First, this feedback strategy is the most used by the professors in the majors where these learners are enrolled. Therefore, after four years or more of exposure
to this feedback type, two options are plausible: that students are really convinced about its efficacy in helping them write more accurately or that learners’ preference is mainly influenced by previous error correction practice and not by real awareness of its value. Research has already pointed out that previous learning experiences can influence learners’ preferences (e.g., Schulz, 2001), which further reinforces the need to uncover learners’ feedback preferences. When elicited, the reported preferences may not be exactly what is best for learners (Ferris, 2004; James, 1998) due to lack of knowledge about effective feedback practices (even about what corrective feedback is). Thus, echoing previous research (e.g., Diab, 2005a; Leki, 1991; Saito, 1994; Schulz, 1996, 2001), L2 teachers might want to have a classroom discussion to bridge the gap between what (low and high proficiency) students believe could be effective and what really is. After all, preference is not indicative of effectiveness (James, 1998), and although there are feedback strategies for the whole gamut of feedback preferences, catering to every learner’s needs and expectations is neither realistic nor feasible.

Second, the significant difference in error correction techniques can also be explained by looking into how much involvement LP and HP learners want in correcting the errors themselves. For instance, the results showed that LP learners’ preferred having the errors located and corrected for them (66.7%) more than HP students (44%). These reported preferences reflect that HP students may feel more linguistically ready to work out the correction themselves as opposed to their LP counterparts, who could lack the linguistic confidence to depend less on the teacher (e.g., “The feedback did help me to correct my mistakes, but what if my correction was also wrong”). As for learners’ preferences regarding amount of correction, while there had already been plenty of evidence that learners prefer comprehensive feedback (Diab, 2005b; Hamouda, 2011; Lee, 2005; Leki, 1991; McMartin-Miller, 2014), our study makes a contribution by showing that such desire is stronger for students in lower proficiency levels. Thus, our findings not only provide support to Oladejo’s (1993) assumption “that with increasing levels of competence, the preferences and wants of learners will change” (p. 74) but also raise the question of whether a one-size-fits all approach to feedback amount is pedagogically effective irrespective of learners’ L2 proficiency level.

Finally, a percentage of HP students in our study preferred being as meticulously corrected as their LP counterparts (cf. Appendix E), which L2 composition teachers could find alarming from a self-efficacy point of view. That is, HP students’ preferences may speak of their self-perception as independent learners even after four years of studying English. Lee (2008) found a similar “tendency for the students to wish for more” (p. 151) and linked it to the teachers’ pedagogical approach to feedback, which tended to be more teacher-centered and less inviting to problem solving and self-correction. This explanation could also apply to the
Comprehensive corrective feedback on low and high proficiency writers

Therefore, as Ferris (2004) suggests, learners need to be given opportunities to develop their self-editing skills by letting them engage in feedback that promotes cognitive problem solving, except for low level students. She states that low proficiency learners lack the linguistic competence to self-correct, and from the results in our study, it is a possibility that they may lack the linguistic confidence as well. This implies that opportunities for self-editing skills should be gradually granted as learners’ developmental level progresses. Besides, applied to high proficiency learners, Ferris and Roberts’ (2001) advice that teachers need to able to explain their feedback choices to learners when they provide implicit marking techniques to students who might not be linguistically ready, also holds for our investigation: L2 teachers should be able to justify why they use highly explicit CF techniques when learners may be linguistically ready to exercise learner autonomy. If advanced learners do not feel that way, it calls for reflection on the part of L2 teachers. As Belcher and Liu (2004) point out, unless teachers relinquish some control over their feedback practices, learners will likely abdicate their responsibility for their own learning.

Conclusion

This study was a first step towards enhancing our understanding of the effects of comprehensive direct CF and MR on learners’ immediate grammatical accuracy and grammatical improvement in light of their proficiency level. It also sought to gain more insight in the relation between proficiency level and learners’ feedback preferences and attitudes. However, this investigation was conducted with low proficiency and high proficiency students within a learning-to-write and writing-to-learn FL university setting, so the findings might not be representative of all learning contexts and learners.

Finally, although we believe the significance of our study lies in its contribution to both SLA and L2 writing fields, the following limitations and qualifications are important to consider for a future research agenda. First, even though the context where this investigation took place provides a much needed ecological validity in feedback studies, caution must be exercised regarding the generalizability of our findings due to the size of our general sample and of our high proficiency learners in particular. Hence, in order to verify our findings, studies with a larger sample size await empirical investigation. Second, the need for a more longitudinal study remains in order to determine whether or not the feedback effect can dilute in the longer term. Third, our study did not look into the effects of the feedback on
individual linguistic features and their amenability to correction. In this sense, echoing Bitchener and Ferris (2012) and Bitchener and Storch (2016), more studies are necessary to investigate to what extent individual targeted features benefit from CF in text revisions and new writings. Thus far, most tightly controlled studies have mainly targeted from one to three features (e.g., Bitchener & Knoch, 2010a, 2010b; Diab, 2015), yet they have found that error amenability to CF may depend more on the error type than on the feedback technique. Fourth, our questionnaire only elicited learners’ attitudes and reasons behind their answers. It did not probe where students’ favorable attitude towards feedback may come from. For instance, Schulz (2001) listed three possible factors that could influence students’ positive attitudes towards corrective feedback (and grammar instruction): previous FL teaching and testing methods, entrenched beliefs about usefulness of error correction, or true convictions shaped by previous learning experiences. Therefore, future work should address the underlying reasons behind learners’ attitudes towards the feedback received. Fifth, in our desire to ensure anonymity and not inhibit learners’ response, our questionnaire to elicit learners’ attitudes was anonymous, so participants did not indicate what treatment they received. This allowed us to establish attitudinal comparisons between proficiency groups. Nevertheless, as informative as level comparisons are, comparisons across conditions and the triangulation of data (process and product data) are deemed desirable to further understand learners’ attitudinal engagement after feedback provision. Sixth, we argued that learners’ feedback preferences might have been influenced by years of previous exposure to certain feedback types within their particular instructional context. However, in order to substantiate our interpretation, future studies need to probe not only learners’ preferences but also the rationale behind them. Lastly, the feedback strategies employed in this study matched more the preferred feedback types and amount of lower proficiency students. Thus, it seems plausible that for this reason their reaction to feedback seemed more welcoming than that of more proficient ones. Conversely, the treatment deviated more (albeit not strongly) from high-proficiency learners’ self-reported preferences, which could have influenced their attitudinal engagement. The implication that an alignment (or lack thereof) between teachers’ feedback practices and learners’ feedback preferences can affect learners’ attitudes towards feedback is noteworthy because it could complement research evidence that indicates that there may not be feedback uptake when learners resist feedback that contradicts their beliefs (e.g., Storch & Wigglesworth, 2010; Swain & Lapkin, 2002, Swain, 2006). Nevertheless, further studies will need to be undertaken to determine exactly to what extent learners’ attitudes and feedback preferences correlate with grammatical accuracy.
References


Bruton, A. (2009). Improving accuracy is not the only reason for writing, and even if it were…. *System*, 37(4), 600–613. doi: 10.1016/j.system.2009.09.005


Comprehensive corrective feedback on low and high proficiency writers


---

**Appendix A. Sample writing prompt**

Read the short excerpt about Facebook. Then, follow the instructions below.

Facebook: Is the online social networking service 1fading?  
New York Times Nov 17, 2013. 07:46PM IST  
Just a few years ago, most of my online social activity 2revolved around Facebook. But lately, my 3formerly hyperactive Facebook life has 4slowed to a crawl. Is it just me, or is Facebook fading? This month David A. Ebersman, Facebook’s chief financial officer 5acknowledged “a decrease in daily users, specifically among younger teens.” The comment confirmed what many of us had suspected but were never able to prove – that the service had become less 6appealing for at least some of its users.

Notes  
1. disappearing  
2. centered  
3. previously  
4. reduced significantly  
5. admitted  
6. attractive

*Article written by Jenna Wortham and adapted from* http://articles.economictimes.indiatimes.com/2013-11-17/news/44162135_1_ebersman-facebook-daily-users

Now consider the following questions:

a. Do you agree or disagree with the ideas expressed in the text above? Explain.
b. What is Facebook good or bad for?

**Instructions:** Answer both questions in order. Develop your answers thoroughly in a 300-word composition. Do not forget to use examples from your own experience to support your general ideas.
Appendix B. Questionnaire of attitudinal engagement and feedback preferences

Please answer this instrument as honestly as possible. Remember that the answers are not graded, nor will they influence in any way the evaluation of the course you are currently taking.

PART A. Instructions:

Indicate with a check mark (√) to what extent you agree or disagree with the following statements about the feedback you received. In the blank space below some of the statements, please explain the reason(s) for your answer. Keep the following abbreviations in mind:

Strongly Agree (SA), Agree (A), Disagree (D), and Strongly Disagree (SD)

<table>
<thead>
<tr>
<th>Statement</th>
<th>SA</th>
<th>A</th>
<th>D</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I found the feedback useful.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For what reason(s)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. The feedback helped me to understand what was wrong.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For what reason(s)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. The feedback was clear.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For what reason(s)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. The feedback was necessary to help me improve my grammar.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For what reason(s)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. The feedback motivated me to revise my text.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For what reason(s)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. I feel the feedback was appropriate for my English level.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For what reason(s)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. The feedback helped me to revise my text successfully.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For what reason(s)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. My grammatical accuracy improved in each text revision.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For what reason(s)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. The feedback helped me to correct my errors.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For what reason(s)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. The feedback I received was too easy considering my existing knowl-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>edge of English grammar.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For what reason(s)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. I wish I had been corrected differently.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For what reason(s)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PART B. Instructions:

Below is a list of error correction techniques. Find the one you believe suits you best and circle its number.

1. Putting in the margins of my text as many check marks as there are mistakes in each line
2. Underlining (or circling) my mistakes only
3. Underlining (or circling) my mistakes plus a code above to know their type
4. Underlining (or circling) my mistakes together with a grammar rule or explanation about them
5. Underlining (or circling) my mistakes together with their correct form above
6. Underlining (or circling) my mistakes together with their correct form above plus a grammar rule or explanation about them
7. Indicating my mistakes in the text and giving me a file with examples of their correct usage
8. Giving me a reformulated text of my composition to see a correct native-like version of my structure

PART C. Instructions:

For each set of statements, circle the letter of the sentence that best expresses your opinion.

1. a. I want feedback that tells me where my error is and gives me the correct form.
   b. I want feedback that indicates where my error is but allows me to make the correction myself.
2. a. I need feedback on all my errors.
   b. I don’t need feedback on all my errors.

Appendix C. Semi-structured interview

1. Were there any corrections that you did not understand? If so, can you show me some examples?
2. What exactly in those corrections do you not understand?
3. Were there any kinds of written comments on your compositions that you wish you could get?
4. What kinds?
5. Do you feel the feedback you received was appropriate for your English level?
6. Do you feel that your instructor’s corrections help you to improve your composition writing skills? Why or why not?
7. What aspects of writing?
8. What do the teachers’ corrections help you improve then?
9. Is there any other kind of feedback you wish you could get on your papers in your major? From peers? From a computer? Both? Why?
10. What do you think has helped you the most to improve your English writing in the major?
Appendix D. Summary of interview answers

Summary of verbatim answers by theme, experimental condition and level

<table>
<thead>
<tr>
<th>Theme</th>
<th>LP</th>
<th>MR</th>
<th>HP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback quality</td>
<td>Interesting Clear</td>
<td>Very useful because it is explained, so I learn.</td>
<td>I actually liked it. Clear</td>
</tr>
<tr>
<td>Feedback benefit</td>
<td>It didn’t help that much.</td>
<td>I could see the errors and correct them</td>
<td>It works when we write fast, and we don’t have time to make revisions.</td>
</tr>
<tr>
<td>Feedback appropriateness</td>
<td>It was O.K.</td>
<td>It was appropriate for my level Content</td>
<td>Ir was appropriate.</td>
</tr>
<tr>
<td>Learners’ wants</td>
<td>Corrections with rules Organization and content</td>
<td>Content</td>
<td>Content</td>
</tr>
<tr>
<td>Writing self-rating</td>
<td>Between poor and good</td>
<td>Between poor and good</td>
<td>Good</td>
</tr>
<tr>
<td>Error detection ability</td>
<td>Good</td>
<td>Poor</td>
<td>Very good</td>
</tr>
<tr>
<td>Views on written CF</td>
<td>It helps a lot because they help us see our own mistakes.</td>
<td>It creates awareness.</td>
<td>It makes you aware of your mistakes.</td>
</tr>
</tbody>
</table>

*Note. LP = low proficiency learners; HP = high proficiency learners; DCF = direct corrective feedback; MR = metalinguistic reminders.*
Appendix E. Summary of Chi-square test for preferred feedback technique

<table>
<thead>
<tr>
<th>Preferences</th>
<th>LP</th>
<th>HP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 (100%)</td>
<td>1 (100%)</td>
</tr>
<tr>
<td>Marginal CF</td>
<td>44 (100%)</td>
<td>0 (100%)</td>
</tr>
<tr>
<td>Indirect CF with underlining (or circling)</td>
<td>41 (93.2%)</td>
<td>3 (6.8%)</td>
</tr>
<tr>
<td>Metalinguistic CF with codes</td>
<td>40 (90.9%)</td>
<td>4 (9.1%)</td>
</tr>
<tr>
<td>Metalinguistic CF with rule reminders</td>
<td>33 (75%)</td>
<td>11 (25%)</td>
</tr>
<tr>
<td>Direct CF</td>
<td>36 (81.8%)</td>
<td>8 (18.2%)</td>
</tr>
<tr>
<td>Direct CF plus codes or rule reminders</td>
<td>33 (75%)</td>
<td>11 (25%)</td>
</tr>
<tr>
<td>Electronic CF</td>
<td>42 (95.5%)</td>
<td>2 (4.5%)</td>
</tr>
<tr>
<td>Reformulation</td>
<td>39 (88.6%)</td>
<td>5 (11.4%)</td>
</tr>
</tbody>
</table>

Note. LP = low proficiency learners; HP = high proficiency learners.

Appendix F. Summary of Chi-square test for preferred feedback degree and explicitness amount

<table>
<thead>
<tr>
<th>Preferences</th>
<th>LP</th>
<th>HP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 (100%)</td>
<td>1 (100%)</td>
</tr>
<tr>
<td>I want feedback that tells me where my error is and gives me the correct form.</td>
<td>16 (33.3%)</td>
<td>32 (66.7%)</td>
</tr>
<tr>
<td>I want feedback that indicates where my error is but allows me to make the correction myself.</td>
<td>32 (66.7%)</td>
<td>16 (33.3%)</td>
</tr>
<tr>
<td>I need feedback on all my errors.</td>
<td>3 (6.3%)</td>
<td>45 (93.8%)</td>
</tr>
<tr>
<td>I don't need feedback on all my errors.</td>
<td>45 (93.8%)</td>
<td>3 (6.3%)</td>
</tr>
</tbody>
</table>

Note. LP = low proficiency learners; HP = high proficiency learners.
Authors’ addresses

Marisela Bonilla López
Faculty of Arts, KU Leuven
Blijde-Inkomststraat 21 – box 3308
3000 Leuven, Belgium
Belgium
Marisela.bonillalopez@kuleuven.be

Elke Van Steendam
Faculty of Arts, KU Leuven
Campus Brussels, Warmoesberg 26
1000 Brussels
Belgium
elke.vansteendam@kuleuven.be

Kris Buyse
Specific Teacher Training Programme, Faculty of Arts, Leuven / Leuven Language Institute (ILT) / Linguistics Research Unit, Campus Sint-Andries Antwerp
Blijde-Inkomststraat 21 – box 3303
3000 Leuven
Belgium
kris.buyse@kuleuven.be