The Immediate and Delayed Effects of Noticing-Supported Error Correction on Learners' Writing Accuracy

Maria-Elena Solares-Altamirano

Abstract

For nearly twenty-five years, researchers have tried to refute Truscott's claim for the inefficiency of written corrective feedback (WCF) for grammar improvement. Despite numerous attempts, the role WCF plays in grammar learning is still unresolved. Problems with early studies trying to negate Truscott's claim included methodological and ethical flaws that recent studies have overcome. Researchers are, however, still unable to totally refute Truscott's claim. This research outlines possible reasons for this. These reasons are presented with the support of a quasi-experimental study on the effects of noticing-supported error correction conditions (ECCs) on learners' writing accuracy. The research tackles comprehensive WCF in an open, uncontrolled and learner-centred writing task. Learner-initiated noticing opportunities were encouraged in each ECC during the feedback stage with the aim to investigate whether learner-initiated noticing opportunities added to ECCs influence writing accuracy. Regardless of numerous studies on WCF, little research has paid attention to noticing while receiving feedback. The study is framed within the 'noticing' debate, the 'language learning potential of writing' and the 'writing-to-learn' and 'feedback-for-acquisition' dimensions. Results are discussed under the main findings: the lack of statistically significant results for grammatical features that might lend support to Truscott's claim for the inefficacy of WCF for grammatical features, and; the statistically significant results found for the CONTROL- and the SELF- ECCs. The former might support Truscott's claim that learners' time and effort would be more productively spent on writing practice. The latter, might suggest that accuracy improvement might not depend on the WCF technique but on noticing opportunities.

Key words: Written corrective feedback, error correction, comprehensive/unfocused error correction, learner-initiated noticing.

Introduction

The topic of written corrective feedback (WCF) gained prominence with Truscott's (1996) questioning of its effectiveness and a call for its eradication. His claim is, however, supported by careful meta-analyses and qualitative scrutiny of existing empirical studies in WCF (see Truscott, 2007 for details). Truscott claims that L2 learning is not a transfer of information as correction implies; it is unrealistic to offer WCF at the exact point when a learner is ready to acquire a specific structure; it is difficult for one type of WCF to treat different language areas (syntax, morphology, lexis); teachers' lack of competence to give reliable feedback and students' ability and determination to use it effectively are uncertain. Truscott's disbelief in teachers' capability to give reliable feedback does not undervalue teachers' ability to correct but rather emphasizes the complexity of error correction (EC), e.g. a failure to perceive every single error is natural. Proofreading shows this is real, even among native speakers. Teachers may also fail to provide correct forms because, despite knowing the errors, sometimes not even experts have a clear understanding of what causes

Maria-Elena Solares-Altamirano, Universidad Nacional Autónoma de México, solares@unam.mx

them or of what the correct usage is. In WCF, this leads to inconsistency, which may confuse learners and be detrimental for learning as it takes time away from other more productive learning activities. WCF may, according to Truscott, maximally result in the development of explicit declarative knowledge rather than implicit procedural knowledge. For nearly twenty-five years, Truscott's claims have prompted endless studies seeking to respond to his assertions and explore the effectiveness of WCF for L2 learning (e.g. Bitchener, 2008; Bruton, 2009; Ferris, 2004; Guénette, 2007 to mention some). Despite numerous attempts, the role WCF plays in grammar learning is still unresolved. Problems with early studies trying to negate Truscott's claims included methodological and ethical flaws (e.g. the inclusion of a pilot group where participants received no feedback in any of their writings) that recent studies have overcome. Researchers are, however, still unable to totally refute Truscott's claim. This research outlines possible reasons for this. These reasons are presented with the support of a quasi-experimental study on the effects of noticing-supported error correction conditions (ECCs) on learners' writing accuracy. First, the type of feedback and the context of WCF that Truscott refers to have been taken for granted. Truscott's claims refer to real classroom contexts whose variables are numerous and difficult to approach in experimental research. Researchers however, have chosen the type of WCF and grammar structure(s) that are easier to control and to measure. This has resulted in copious investigations:

- on various grammatical structures: usually one, at most three (Bitchener, 2008; Sheen, 2007);
- testing various corrective feedback (CF) techniques: direct vs. indirect (Chandler, 2003; Ferris & Roberts, 2001); with vs. without metalinguistic explanations (Bitchener, 2008; Bitchener, Young & Cameron, 2005), with vs. without tutorial meetings;
- within various contexts: mostly with ESL adult L2 learners;
- with various artificial writing tasks: guided letters (Bitchener et al., 2005); story sequences (Swain & Lapkin, 1995, 2007; Yang & Zhang, 2010) picture descriptions (Hanaoka & Izumi, 2012).

The variations above have caused research results to be fragmented, not generalizable and limited.

Second, the above situation derives from the following requirements of experimental research:

i. Controlled environments imply isolating the writing process from other classroom variables, such as multiple drafting, whole-class feedback and instruction.
ii. Controlled variables have led to a focus on: (a) specific grammar targets, with the English article system being the most frequently analysed structure (Bitchener & Knoch, 2008, 2009a, 2009b); (b) specific language proficiency levels, mostly intermediate and advanced; (c) a reduced number of words e.g. 125, 250 words, ten-line letters.
iii. Controlled tasks have also translated into unauthentic controlled writing tasks: picture descriptions (Qi & Lapkin, 2001); picture sequence narratives (Swain & Lapkin, 1995, 2007; Yang & Zhang, 2010); or guided letters (Bitchener et al.,

2005) where writing is approached as a means to practise grammar rather than an end itself).

iv. Controlled procedures have also translated into pre-, post- and delayed post-test sequences.

As such, the demands of experimental research have, to a certain extent, reduced the ecological validity of research designs, making them distant from classroom realities.

Third, published WCF research seems to be inclined towards studies supporting WCF (Truscott, 2007). The Ellis, Sheen, Murakami, & Takashima, H. (2008) study, for instance, is presented as dealing with comprehensive (also called unfocused) EC, despite the authors themselves, making it clear they did not use comprehensive EC: "In retrospect, it might be better to characterize the differences between the two types of CF in this study as *focused* versus *less focused* rather than *focused* versus *unfocused*" (p. 367). Van Beuningen's (2012) support for the efficacy of comprehensive EC in accuracy improvement is also generalised, despite the uncommon naturalistic teaching context of her study "It also needs to be noted that the present study was performed within a particular context, which could be described as a relatively naturalistic SLA environment" (p. 34). In Truscott's (1996) meta-analysis of WCF studies, he stated published work had been unfair in favouring the supportive effects of correction and offering little space for different views. Bias in favour of correction research makes those studies "look better than they actually are" (Truscott, 2007, p. 267).

The context above added to the author's belief that learners are the ones who decide what, when and how much attention they pay to the feedback they receive while writing led the researcher to design a more pedagogically oriented study on the effects of different noticing-supported ECCs. The study attempts to tackle comprehensive (with no specific linguistic target) EC in an open, uncontrolled, learner-centred, syllabus-based writing task. The nature of learner-initiated noticing (with no predetermined target and depending on learners' willingness to respond to feedback), the characteristics of uncontrolled open writing tasks, and the complexity of comprehensive EC made this research full of challenges and discoveries. This study is part of the author's unpublished PhD thesis (2016). Only one of the two research questions is approached in this paper. The learners' response to noticing opportunities will be dealt with in a different paper.

Theoretical Framework

This study is framed, on one side, within the 'language learning potential of writing' (LLPW), the writing-to-learn and the feedback-for-acquisition dimensions (Manchón, 2011a, 2011b) and; on the other, within the noticing-awareness debate. The latter is only introduced in this paper and will be dealt with in a future publication. Manchón (2011b) puts forward the LLPW as a new research domain, the aim being "to investigate the writing-to-learn language dimension of L2 writing development and instruction" (p. 62). Manchón (2011b, 2013) also differentiates between two areas of investigation: the *learning-to-write* dimension (referring to how people learn to express themselves in writing) focuses on L2 writing research and the *writing-to-learn* dimension (referring to how people's engagement with writing contributes to L2 learning) focuses on SLA studies. Within the learning-to-write dimension, writing constitutes an end in itself and teaching is associated with multi-(literacy) education, i.e. literacy in more than one

language (Manchón, 2013). Within the writing-to-learn dimension, writing is considered a 'means' for language learning and teaching is associated with foreign language instruction. The writing-to-learn a language dimension is concerned with *feedback-for-acquisition* rather than with *feedback-for-accuracy*. The former refers to how writing − text production and feedback processing − fosters L2 development; learners' capacity to exploit the knowledge gained from feedback on previously corrected writing in new writing. The latter refers to the revisions learners make to previously corrected writing.

Regarding *noticing* as an attentional process, the role it plays in L2 learning is well accepted by psychologists and SLA researchers. It has also received great support and prompted much investigation (e.g Adams, 2003; Godfroid, Housen & Boers, 2010; Leow, 1997; Mackey, 2006). However, the close relationship between noticing, consciousness and awareness, different theorists' (Leow, 1997,; Robinson, 1995; Schmidt, 1990, 1995, 2001; Tomlin & Villa, 1994) understanding of these terms and disagreement on the role of consciousness in L2 learning have turned this topic into one of the most controversial debates in SLA. The importance of attention in L2 learning is unquestioned among theorists. However, Reindeers (2005) points out that "[Theorists] differ greatly in how they explain the storage and retrieval of information" (p. 31). It is agreed that attention is necessary for learning; disagreement emerges over whether attention involves awareness or not.

The Present Study

In this study, the effects of noticing-supported ECCs at the feedback stage on learners' writing accuracy in the rewriting and new writing stages are explored. In most WCF studies, noticing has been assumed to occur spontaneously after feedback delivery. The researcher's hypothesis is that input provided by different types of WCF must be accompanied by noticing opportunities to ensure that the learner will direct their attention to the input provided. The amount and error types that learners noticed correspond to the second research question of the study which is not approached here. This paper focuses on one research question only: the effects of noticed-supported ECC on accuracy improvement. Given the abundant studies in WCF, why do more research on this topic? Despite the numerous worldwide studies on WCF, many of them have been performed in ESL and immersion contexts. Mexico offers an EFL context with scarce research in this topic. Furthermore, as explained in the previous sections, current research still has limitations: it is mainly non-comprehensive, it generally includes only one piece of writing, tasks are too controlled and generally short (200 words maximum), feedback treatments are not sustained, studies are performed in controlled environments. Research designs have also disregarded learners' engagement with feedback, the role of practice and the processing of feedback. Many of these limitations are explained by the demands of experimental research. Storch (2010) remarks that "in the desire to conduct more robust research, the pendulum has swung too far towards experimental studies" (p. 29). The priority in WCF studies has been on testing the effectiveness of different types of teacher-provided feedback, and little attention has been paid to *learner-initiated noticing*, i.e. what learners 'notice' or 'attend to' by themselves while receiving feedback. Storch and Wigglesworth (2010) explain that noticing and the "processing of feedback [are] … less … researched and understood because it is difficult to access such learner-internal

cognitive processes" (p. 305). However, Santos, López-Serrano, and Manchón (2010 p. 32) indicate that because of the self-initiated character of writing problems, writing becomes the perfect setting to study self-initiated noticing and focus-on-form (FonF) processes. With the desire to contribute to the research on WCF, an effort was made with this study to address some of the previously mentioned limitations. Above all, an effort was made to plan a more ecological and classroom representative design**.** First, the writing task used in this study was an open, uncontrolled, comprehensive, learner-centred, 300-word opinion essay. Contrary to other studies where the task is selected according to the researcher interests, the ability to write opinion essays is a syllabus-based task for the target students. A 300-word opinion essay is also longer than tasks in previous studies. Second, despite Manchón's (2011b) call for more investigation on feedback for acquisition, no previous study on noticing in WCF (to the best of the researcher's knowledge) has included *a new writing task*, i.e. analysis going beyond revision. Existing studies on noticing in WCF have only worked with composing and revision tasks (Qi & Lapkin, 2001; Santos et al., 2010; Yang & Zhang, 2010). Third, some WCF research has been performed with collaborative writing. This design explores individual writing. "Given that many forms of writing are intrinsically an individual enterprise, research findings on collaborative writing should not be taken to represent potential learning benefits of writing per se" (Manchón, 2011b, p. 76). Fourth, this research joins the few studies attempting to approach *comprehensive* (correction at all levels) *WCF*, the most time-demanding and frequently used WCF technique in FL classrooms. Research evidence opposing Truscott comes mostly from non-comprehensive WCF studies using controlled writing tasks. As long as evidence for the benefits of WCF comes from non-comprehensive rather than comprehensive experimental studies, it will be difficult to refute Truscott's claims (Ellis et al., 2008). Fifth, different from previous studies, the treatment in this study adds noticing opportunities to the ECCs tested, i.e. the treatment consisted of 'noticing-supported error ECCs or WCF'. This treatment aimed to explore the potential of 'noticing-supported ECCs' for learners' writing accuracy. The researcher believes that despite the importance of noticing and attention for learning, learners' willingness to respond to feedback is usually presupposed. The researcher suggests that providing WCF does not automatically imply that learners will pay attention. First, noticing opportunities have to be provided and attention to feedback has to be confirmed. Previous studies have looked at teacher-prompted noticing (Santos et al., 2010). The researcher gives priority to noticing which is learner-generated (Hanaoka, 2007; Park, 2011; Williams, 2001), noticing which is not influenced by teacher intervention. Finally, the study extends the range of acquisition contexts by including samples of students in EFL contexts. Van Beuningen's (2011) research greatly shaped this design. Her study encouraged the error categorization (grammatical vs. non-grammatical) and quantitative direction of the study. Grammatical errors included syntactic and word form errors (e.g. verb tense, singular/plural), word order, incomplete sentences, and addition or omission of a word. Non-grammatical errors included lexical, appropriateness / pragmatic errors, spelling, punctuation, and capitalization errors. The selected ECCs in this study are:

Direct error correction: A WCF technique where all error types are signalled and its correction presented. It was selected because, despite its disadvantages (isolated corrections, unclear comments, lack of learner's engagement with cognitive processing and emphasis on faults), it is a widely used WCF technique in foreign language (FL)

classrooms. Evidence for its efficiency is central to pedagogy and justifies further research.

Reformulation: A WCF technique that corrects errors in a text, maintaining its content but offering a native speaker's (or proficient L2 speaker's) version (Cohen, 1983; Johnson, 1988). It was included in this study as it is a learner-centred, tailor-made feedback technique. It is one of the least intrusive WCF techniques and has proved to be effective in promoting learners' noticing (Allwright, Woodley & Allwright, 1988; Yang & Zhang, 2010). Reformulation was operationalised in this study by having a native writer of the target language rewrite the learner's text, maintaining his/ her ideas, making it as native-like as possible (Cohen, 1983, p. 4). Three native speakers rather than one were necessary because of the task length and the number of essays to be reformulated within a short period of time. Reformulators received previous training and participated in previous pilot study.

Self-correction: An ECC that implies no external explicit feedback, i.e. the learner self-corrects his/her production after monitoring their own output. It was selected because it triggered noticing during a previous pilot study. There, learners were able to notice their own errors immediately after they received their original text, i.e. even before feedback was provided. With 'self-initiated noticing' being the focus of this research, it was important to include a self-correction condition.

The difference between *feedback* and *error correction condition* becomes important for this research. *Feedback* (the what) is defined as input (in the form of information) about the correctness (what is acceptable in the L2) or incorrectness (what is not acceptable in the L2) of learners´ linguistic performance, speech or output. Feedback aims to enable learners to correct their inaccuracies.

Learners correct their inaccuracies and modify their output if necessary. Feedback may be: *(a) external* (information is provided by someone else, e.g. the teacher, more advanced interlocutors, L2 native speakers or the environment), or *(b) internal* (information resulting from learners attempts to achieve correctness by themselves, i.e. information is self-provided or self-initiated as learner's self-correction). A variety of WCF techniques exist depending on how feedback is provided: its explicitness (direct or indirect), its focus (focused or unfocused, also known as non-comprehensive and comprehensive respectively) or the person delivering it (teacher feedback, peer feedback, self-correction).

Error correction condition (the how) refers to the specific techniques (explicitness, focus, person delivering it) used to provide feedback in each experimental group. The three ECCs included in this study were direct and comprehensive but differed in terms of the person delivering the corrections. Therefore, ECC was operationalised as internal (learner self-provided) or external (provided by others) information about learners' linguistic performance in their written essays. ECCs differed concerning the person delivering feedback: the researcher in direct error correction (DIR-ECC), native speakers in reformulation (REF-ECC), and learners themselves in self-correction (SELF-ECC). In the REF and DIR groups, feedback was external, i.e. provided by others. In the SELF-correction group, feedback was internal, i.e. provided by learners' themselves, as self-provided or self-initiated feedback, which the researcher defines as information about the correctness or incorrectness of a learner's linguistic performance that derives from the learner's self-correction.

Other important constructs in this study are (a) *learner-initiated noticing*, adapted from Godfroid, Housen, & Boers (2010, p. 169). As "episodes when learners pay attention, by themselves in the absence of any external intervention, to new linguistic data in the input, and relates these to existing knowledge" Noticing and learner-initiated noticing are operationalized as learners' written reports of their language difficulties reported on noticing sheets; (b) *accuracy* operationalized as the percentage of correct usage of grammatical and non-grammatical features (Van Baungheim, 2011).

Given the considerations above, the research question (RQ), and the subsequent queries that guided this study are:

- What are the effects (if any) of different noticing-supported ECCs at the feedback stage on learners' writing accuracy in the rewriting and new writing stages?
- Does the input provided by the above noticing-supported ECCs at the feedback stage have any effects on learners' writing accuracy in the rewriting and new writing stages?
- If so, what error types (grammatical or non-grammatical) are more amenable to correction in different noticing-supported error correction conditions?

Method

Participants and Setting

This study was conducted at the National Autonomous University of Mexico (UNAM) with students in tertiary education. Despite the fact that English proficiency constitutes a graduation requirement for UNAM students, neither credits are gained nor are requirements met by studying at Escuela Nacional de Lengua Lingüística y Traducción (ENALLT). English becomes an extracurricular subject and students attend courses on a voluntary basis. All participants were Spanish speakers whose English level was upper-intermediate (equivalent to B2 level in the Common European Frame of Reference). EFL is taught at ENALLT in four-skill courses. Classes were two hours long and took place three times per week.

The population of this study consisted of 60 students ($N = 60$) in their second year of tertiary education. Participants were divided into four groups of 15 students each. The mean age of participants' total sample was 23.6, ($SD = 4$, Minimum = 19, Maximum = 38). Participants were recruited as they registered for their upper-intermediate EFL term. Experimental and control conditions were randomly assigned to the four groups. Students' participation was both anonymous and voluntary.

Data Analysis

One hundred and eighty 300-word essays produced by four groups in the composing (60), rewriting (60) and new writing (60) stages were analysed. Essay analysis required quantitative, descriptive and inferential analysis. Learners' essays were first coded for linguistic errors. Grammatical and non-grammatical error ratios ([number of linguistic errors/ total number of words] x 100) were computed for different dependent variables (i.e. overall, grammatical, non-grammatical accuracy) in the composing, rewriting and

new writing stages. ANOVA and post-hoc tests for statistically significant results were performed. Essay analyses for selected error types were also performed. Rubrics for error categorizations resulting from the pilot study were used.

The Design

A four-stage (composing/ error correction-noticing/ rewriting/ new writing) writing task including three different ECCs (DIR, REF, SELF) and a control group was designed. Four groups participated in the same writing task and experienced the same treatment: noticing-supported ECC. However ECCs were different for each experimental group: DIR-EC (group 1), REF-EC (group 2) and, SELF-EC (group 3). Group 4 was the control group. The effects of noticing-supported ECCs (independent variable) on learners' writing accuracy (dependent variable) across tests (pre- post- and delayed-post) was measured.

Procedures

All tasks and treatments took place during class periods and were set up by the researcher. The experiemental procedure is summarized in Table 1.

Table 1. Experimental procedure

	Week 1	Week 3			Week 4
	SESSION 1 (Stage 1)	SESSION 2 (Stages 2 and 3)			SESSION 3 (Stage 4)
	Composing Pre-test *Essay 1*	Error correction condition	*Noticing Sheet*	Rewriting Post-test *Essay 2*	New-writing task Delayed post-test *Essay 3*
		Treatment			
Group 1 *N* = 15	✓	✓ Direct error correction	✓	✓	✓
Group 2 *N* = 15	✓	✓ Reformulation	✓	✓	✓
Group 3 *N* = 15	✓	✓ Self-correction	✓	✓	✓
Group 4 *N* = 15	✓	X No error correction	X No noticing sheet	✓	✓
60	60		45	60	60

Class teachers were absent from all experimental sessions. The researcher corrected essays in DIR-G; three previously trained native speakers (who participated in previous pilot study) reformulated the essays in REF-G. To avoid onerous pen-and-paper work, experimental sessions were performed in a computer room especially adapted for this study. Learners typed their essays on computers (grammar and spell-check functions deactivated and no Internet access). Each student had their own folder in their personal computer. After each session, students' work was sent to the researcher's computer via a

server. The design required three two-hour sessions: session one (stage 1), session 2 (stages 2 and 3), session 3 (stage 4).

Session 1: Stage 1 – Composing or Pre-test (Week 1). During this composing stage, the four participant groups completed writing task 1, an opinion essay on a given prompt. CONTROL-G experienced no treatment (no error correction or noticing sheet). Instead, the learners wrote essay 1 (session 1), re-wrote essay 1 (session 2) and wrote a new essay (session 3). CONTROL-G received no WCF other than the researcher's general comments on the content of their writing. Five short story books were raffled among the participants in CONTROL-G to thank them for their participation.

The four groups received a writing prompt with instructions for writing task 1. Learners were informed about task requirements: length, time available and no dictionary support. Students had a two-hour session to write and proofread their essay before submitting it. At the end of the session, essays (in the four participant groups were sent to the researcher's folder. Each essay was printed twice (once for error analysis by the researcher, and once for the students at the noticing stage). Fifteen essays from the DIR-G were corrected by the researcher. Fifteen essays from REF-G were divided into three sets. Each native speaker received a set of essays and had ten working days to reformulate them. Corrections in the DIR and REF groups were made electronically by the researcher and by the native speakers, respectively. Only printed versions were returned to students for their error inspection in the noticing stage.

Session 2 (First Hour): Stage 2 – Error Correction + Noticing (Week 3). The treatment session (session 2) was held 12 days after session 1 to allow time for essay correction and reformulation. During this stage, learners in the DIR and REF groups received their printed original essay and were given time to read it and recall what they had written. DIR-G received their error corrected essay and a noticing sheet whereas REF-G received their typed reformulated essay and the same noticing sheet. The DIR and REF groups compared their feedback with their original writing and completed noticing sheet. This uncontrolled condition aimed to collect evidence about what learners noticed (if they did so) by themselves, as the noticing sheet was an almost blank sheet of paper with no teacher's guidelines on what to look at.

Learners in SELF-G received their printed original essay with no alterations and the same noticing sheet that the DIR and REF groups received. Learners reread their essay, identified their own mistakes and corrected them. Detected corrections were reported on the noticing sheet. Learners in CONTROL-G neither received their printed original essay nor engaged in feedback analysis. Instead, learners rewrote a second essay with the same prompt as in essay 1. Consequently, this group finished one hour before the others. All experimental groups had one hour to analyse their feedback and complete the noticing sheet. CONTROL-G moved directly to rewriting essay 1; learners in all groups had the same amount of time (one hour) for rewriting. Time on task was tested in the pilot study where there was no time limit, one hour proved sufficient for the task.

Session 2 (Second Hour): Stage 3 – Rewriting or Post-Test (Week 3). Stage 3 took place in the second hour of session 2. CONTROL-G did not participate in the second hour of session 2 since they had no treatment. This group worked on stage 3 during the

first hour of session 2. After one hour of noticing in the three experimental groups, all materials (original essay, noticing sheets, corrected/ reformulated essays) were collected. The DIR, REF and SELF- groups received writing task 2 (same writing prompt as for essay 1, now called essay 2). Students rewrote their essay on their computers with no support (reformulated/ corrected essay, noticing sheets or dictionaries). Students did not know about this post-test or the new writing post-test. The word rewriting is used instead of revising as 'revising' may imply having access to received error correction, which was not the case (all corrected and self-corrected essays were previously collected). Rewriting the essay was included as a way to engage learners with the feedback received and test whether there was any immediate improvement in accuracy. At the end of session 2, rewritten essays in the experimental and control groups were sent to the researcher's folder.

Session 3: Stage 4 – New Writing or Delayed Post-Test (Week 4). One week after the rewriting session, all groups (three experimental and one control) wrote a new writing task on a new but similar topic (same topic for all groups). One hour was allowed to do this. Before moving onto the next section, an important change to the research design must be explained. Examining comprehensive EC was the original aim of the study. However, the lack of inter- and intra-reliability in accuracy analyses obliged the researcher to move from comprehensive to semi-comprehensive EC. Learners did receive comprehensive EC during their feedback and noticing opportunities. This was crucial for learner-initiated noticing, which implied all error types (not exclusively predetermined linguistic targets) were corrected. It was the essays' error analysis that was semi-comprehensive, i.e. accuracy was measured by considering only seven selected (the most frequent) linguistic features: spelling and lexis (non-grammatical accuracy), omission of constituent, unnecessary definite article, S+V agreement, 3rd person singular and gerunds (grammatical accuracy).

Results

Essay 1 (overall accuracy) was used as a base to find out whether the four participant groups were comparable. When students wrote essay 1 they were in equal conditions, no group had experienced any treatment. The error rate in each group in essay 1 was calculated. Then, a one-way ANOVA on Test 1 scores showed no significant differences between groups: $F(3, 56) = 0.59$, $MSE = 52.1$, $p = 0.63$. This means the groups were comparable. Thus, any improvement from the post-tests was not as a consequence of prior differences between groups.

Table 2. Descriptive statistics for the overall number of errors across the four groups in each test session

Group	Essay 1 (Pre-test)	Essay 2 (Post-test)	Essay 3 (Delayed post-test)
DIR	30.1 (10.0)	22.0 (7.2)	27.3 (5.8)
REF	31.9 (11.0)	27.4 (12.0)	31.9 (14.3)
SELF	32.0 (6.8)	30.1 (6.9)	28.1 (9.3)
Control	28.1 (9.1)	27.7 (10.2)	26.4 (13.5)

Note. Numbers represent mean number of errors with standard deviations in parentheses. *DIR* = Direct group, *REF* = Reformulation group, *SELF* = Self-correction group, *SD* = Standard Deviation

The four groups' accuracy performance across the three essays in terms of the number of errors for each test session is presented in Table 2. As observed in Table 2 all groups showed overall accuracy improvement from essay 1 to essay 2 (the numbers of errors in the four groups went down: from 30.1 to 22 for DIR; from 31.9 to 27.4 for REF; from 32 to 30.1 for SELF; from 28.1 to 27.7 for CONTROL). However, from essay 2 to essay 3, the number of errors in DIR and REF went up again: from 22 to 27.3 for DIR; from 27.4 to 31.9 for REF. The number of errors in SELF (with self-provided or self-initiated feedback and noticing opportunities) and CONTROL (no feedback, no noticing opportunities) from test 2 to test 3 again went down: from 30.1 to 28.1 for SELF; from 27.7 to 26.4 for CONTROL.

The immediate overall accuracy improvement observed in all participant groups was partially retained until test 3 in the two groups receiving external explicit feedback (DIR and REF groups). Mean error in the DIR and REF groups went up again in test 3. However, the mean error seen in test 3 was still lower than in test 1 (for DIR) or equal to test 1 (for REF). SELF-G (with self-provided or self-initiated feedback and with noticing sheets), however, was the only experimental group that improved its accuracy across the three essays: 32.0 (essay 1), 30.1 (essay 2), 28.1 (essay 3). Concerning CONTROL-G, (no feedback, no noticing opportunities), its performance was similar to that of SELF-G, i.e. its overall accuracy improved across the three tests: 28.1 (essay 1), 27.7 (essay 2), 26.4 (essay 3).

Figure 1 illustrates the groups' overall accuracy performance across the tests. Note that the numbers of errors in the DIR and the REF groups go down from test 1 to test 2, though they go up again from test 2 to test 3. The numbers of errors in the SELF and the CONTROL groups, conversely, go down across the three tests.

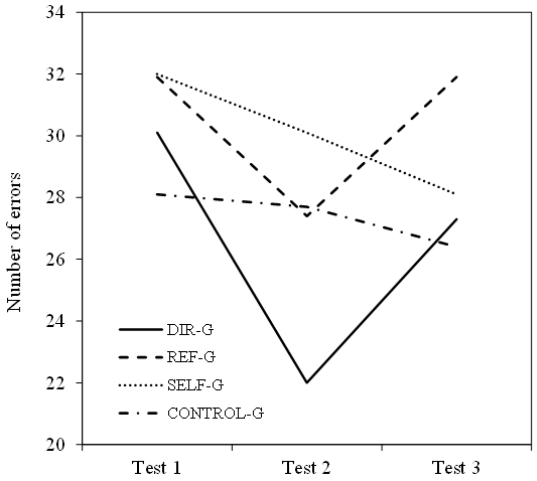

Figure 1. Groups' overall accuracy performance across tests.

Table 3 shows that non-grammatical features mostly determined the trend described above for the overall accuracy performance of the four groups. N-GR errors went down from essay 1 to essay 2 in the DIR (from 18.4 to 11.6) and REF (from 18.0 to 17.5) groups, though N-GR errors went up again in test 3 for DIR (from 11.6 to 13.8). The REF group was the exception as, different from the overall accuracy performance, the number of N-GR errors kept going down: from 17.5 (essay 2) to 16.3 (essay 3). N-GR errors in the SELF and CONTROL groups, as in overall accuracy performance, kept going down across the three tests.

Table 3. N-GR and GR accuracy performance of the four groups across the tests

Group	Essay 1 (Pre-test)		Essay 2 (Post-test)		Essay 3 (Delayed post-test)	
	N-GR	GR	N-GR	GR	N-GR	GR
DIR	18.4 (5.7)	11.7 (6.2)	11.6 (5.2)	10.4 (5.5)	13.8 (6.7)	13.5 (5.0)
REF	18.0 (6.9)	13.9 (6.9)	17.5 (8.0)	9.9 (4.8)	16.3 (9.2)	15.6 (7.8)
SELF	19.8 (4.5)	12.2 (5.5)	17.7 (4.9)	12.5 (6.7)	15.7 (3.8)	12.3 (8.3)
Control	15.7 (5.9)	12.4 (6.4)	16.2 (6.9)	11.5 (5.1)	12.6 (5.6)	13.8 (8.9)

Note. Numbers represent mean number of errors with standard deviations in parentheses. N-GR = Non grammatical, GR = Grammatical, SD = Standard deviation, DIR = Direct group, REF = Reformulation group, SELF = Self-correction group.

Inferential analyses were performed to determine if the results in Table 2 and Table 3 were statistically significant, i.e. if the results were not attributed to chance.

Before that, the researcher confirmed that data met the assumptions[1] of an ANOVA. Figure 2 presents an example of normal distributions for the direct group. Similar distributions were found in the other three groups. Confidence intervals were calculated with $\alpha \leq 0.05$.

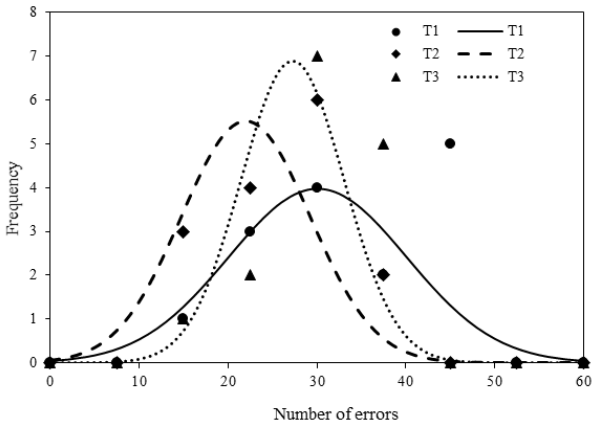

Figure 2. Normal distributions for overall accuracy of DIR group across tests. Similar distributions were found for the other three groups.

A number of within groups and between groups one-way ANOVA tests were performed to measure the effects of different noticing-supported ECCs on learners' language accuracy within groups[2], i.e., whether each group's accuracy performance improved, or not, across the three tests and between them, i.e., comparing the four groups' accuracy performance in each test, (whether groups continued to be comparable or not as they were in test 1). Thus, the inferential statistics were conducted in order to answer whether there were any differences in the overall, non-grammatical and grammatical accuracy performance of the DIR, REF, SELF and CONTROL groups across the three tests.

Table 4 presents the results of within groups ANOVA tests for overall accuracy and non-grammatical errors. Grammatical category is not included in the table as there were no statistical differences, i.e. it did not improve.. The analyses revealed significant results in the three tests for overall (p = 0.03) and non-grammatical (p = 0.01) accuracy in the DIR group, suggesting that the latter influenced the former. Significant results were also found for non-grammatical accuracy in SELF-G (p = 0.05). Post hoc

[1] (1) Groups must be independent of one another; the same data must not be contained in two groups; (2) the residuals (differences from the mean) must be approximately normally distributed; (3) the residuals must have approximately equal variances.

[2] The word 'groups' in 'within and between groups' does not refer to the four participant groups in the study. 'Groups' in ANOVA terminology refers to the performance of each of the four participant groups: (a) Across the three tests when it is read 'within groups', i.e. whether they improve or not; (b) In each of the tests when it is read 'between groups', i.e. whether they are comparable or not, as in test 1.

Cohen's d tests suggest that the effect sizes for the significant results were small.[3] The analyses did not indicated significant results for the REF (p = 0.53), SELF (p = 0.40), and Control (p = 0.92) groups, suggesting that these conditions had no significant effects on (or led to no improvement in) learners' overall accuracy performance across the three tests.

Table 4. Within groups ANOVA results

Group	Overall				Non-grammatical			
	F (2,42)	*MSE*	*p*	*Cohen's d*	*F* (2,42)	*MSE*	*P*	*Cohen's d*
DIR	4.05	251.6	0.03	0.11	5.25	180.6	0.01	0.12
REF	0.65	102.8	0.53	- - -	0.17	11.1	0.86	- - -
SELF	0.96	58.1	0.40	- - -	3.16	62.1	0.05	0.10
Control	0.09	11.4	0.92	- - -	1.50	56.6	0.25	- - -

Note. MSE = mean-square error, *DIR* = Direct group, *REF* = Reformulation group, *SELF* = Self-correction group.

Post-hoc analyses were conducted on the significant effects through t-tests in order to find out whether the significant test differences found in DIR-G (overall and non-grammatical accuracy) and SELF-G (non-grammatical accuracy) lay (a) among all of the tests and between each other or (b) only in one of them.

For the DIR group's overall accuracy results, t-tests[4] (Table 5) showed test differences lay in test 2. The results of tests 1 and 3 were the same and these equal results were different from the results of test 2. The results of test 2 were better (lowest error mean 22) than the results of tests 1 (mean error 30) and 3 (mean error 27).

Table 5. Test comparisons in direct group (overall accuracy)

Test comparison (error mean)	*F*(1, 28)	*MSE*	*p*
(30) T1 – T2 (22)	6.39	488.0	0.02
(30) T1 – T3 (27)	0.88	58.8	0.36
(22) T2 – T3 (27)	4.86	208.0	0.04

Note. MSE = mean-square error

For the DIR-G non-grammatical accuracy results, post-hoc analyses (Table 6) showed test differences lay in test 1. The results of tests 2 and 3 were the same and these equal results were different from the results of test 1. The results of test 1 were worse (highest error mean 18) than the results of tests 2 (mean error 12) and 3 (mean error 14).

[3] Cohen suggests that *d = 0.2* is considered a 'small' effect size, 0.5 represents a 'medium' effect size and 0.8 a 'large' effect size.

[4] A t-test is considered to be a special case of one-way ANOVA. Whereas a t-test is limited to comparing the means of two groups, one-way ANOVA can compare more than two groups.

Table 6. Test comparisons in direct group (non-grammatical accuracy)

Test comparisons (error mean)	*F*(1, 28)	*MSE*	*p*
(18) T1 – T2 (12)	11.77	346.8	0.03
(18) T1 – T3 (14)	4.15	158.7	0.05
(12) T2 – T3 (14)	1.02	36.3	0.32

Note. MSE = mean-square error

For the SELF-G non-grammatical accuracy results (Table 7), t-tests showed that despite a slight performance improvement being observed from test 1 to test 2 (mean error decreased from 20 to 18) and from test 2 to test 3 (mean error decreased from 18 to 16), these differences were not statistically significant. However, a statistical difference emerged between tests 1 and 3 (mean error decreased from 20 to 16).

Table 7. Tests comparisons in self- group (non-grammatical accuracy)

Tests comparisons (error mean)	*F*(1,28)	*MSE*	*p*
(20) T1 – T2 (18)	1.53	34.1	0.23
T1 (20) – T3 (16)	7.13	124.0	0.01
T2 (18) – T3 (16)	1.46	28.0	0.24

Note. MSE = mean-square error

ANOVA tests showed there were no significant differences in the grammatical accuracy performance of any of the participant groups. Results for all groups in the three tests were statistically the same. Thus, no further significance tests were necessary.

Analyses on Selected Error Types

ANOVA tests were also run on the seven selected error types across the three tests. Table 8 shows the accuracy performance of the four groups for different error types, improved exclusively for spelling in the DIR and CONTROL groups. Spelling was the only error type that showed statistical significance within groups, i.e. the only error type with improved accuracy across the three tests. This means error type did not influence accuracy performance (except for spelling in the DIR and CONTROL groups).

Table 8. Error types showing significance within groups

		Spelling	
Group	*F*(2, 42)	*MSE*	*p*
DIR	3.48	76.8	0.05
REF	0.02	0.6	0.99
SELF	0.80	13.1	0.47
control	3.86	59.5	0.04

Note. MSE = mean-square error, *DIR* = Direct group, *REF* = Reformulation group, *SELF* = Self-correction group

Regarding statistical significance between groups, Table 9 shows that the error types (of the seven selected for this study) that showed statistical significance were: non-grammatical accuracy of T2 for the DIR group (p = 0.04); spelling accuracy of T3 for the control group (p = 0.02); lexis accuracy of T2 for the DIR group (p = 0.04).

Table 9. Error types showing statistical significance between groups

Test	Non-grammar			Spelling			Lexis		
	F (3,56)	*MSE*	*p*	*F* (3,56)	*MSE*	*p*	*F* (3,56)	*MSE*	*p*
T1	1.31	44.2	.29	1.07	21.2	.38	1.49	16.2	.24
T2 (DIR)	2.98	121.4	.04	1.79	38.9	.17	2.98	47.5	.04
T3 (Control)	1.02	44.6	.40	3.91	83.0	.02	0.95	17.2	.43

Note. MSE = mean-square error, *T* = Test, *DIR* = Direct group. Only error types with significant results are included.

To summarize, the question the researcher wanted to answer with between groups analyses was whether there were any differences in the overall, N-GR, GR and error type accuracy performance of the DIR, REF, SELF and CONTROL groups in each of the three tests (i.e. whether the participant groups continue to be comparable / equal as they were in test 1 or whether treatment has resulted in a change or made the groups different; if so, which group(s) and in which test(s)?). The researcher already knew, because he/she compared the four groups before treatment, that, in Test 1, the four participant groups were comparable because no statistical differences were found among them before treatment. What the researcher discovered with the between tests analyses for T2 and T3 was that the participant groups were not comparable anymore in Test 2 and Test 3 because treatment did result in statistically significant differences.

The results in Table 9 show that the DIR group performed better than the other participant groups in test 2, especially in lexis, with its best performance in non-grammatical accuracy. These results confirmed the significant differences already found for the overall performance of the DIR group. They also show that the CONTROL group performed better than the other participant groups in spelling in test 3. Although descriptive statistics showed a positive impact from all noticing-supported ECCs on overall accuracy improvement in the rewriting and to a lesser extent the new writing stages, statistical significance did not support these outcomes. Statistical significance across the three tests was exclusive to: (a) overall and non-grammatical accuracy in noticing-supported DIR-ECC and (b) non-grammatical accuracy in noticing-supported SELF-ECC. Concerning the error types included in the accuracy analyses, a series of ANOVAs revealed that except for spelling in the DIR and CONTROL groups, the error type did not influence accuracy performance.

Discussion

Findings are first discussed by answering the research question**.** The results from descriptive statistics revealed that noticing-supported ECCs in the feedback stage had positive effects on learners' writing accuracy in the rewriting and new writing stages. All groups showed overall accuracy improvement from essay 1 (composing) to essay 2 (rewriting). The reason for this immediate accuracy improvement might be attributed to the treatment (error correction and noticing opportunities). Whether it was the error condition, the noticing opportunities or a combination of both, what led to improvement could not be determined at this stage. These descriptive statistics outcomes support findings in previous studies regarding the positive effects of different ECCs in a post- or immediate writing task, i.e. WCF as a revising tool (e.g. Ashwell, 2000; Fathman & Whalley, 1990; Ferris & Roberts, 2001). What is new in this study is that results come from a less common task type: an open uncontrolled learner-produced task, with semi-comprehensive error correction. DIR, REF and SELF ECCs showed positive effects in an immediate test. However, Santos et al. (2010) suggest immediate improvement cannot be considered acquisition. In fact, it is possible that recalling played a role in immediate improvement. Except for the control group, learners in all the ECCs were engaged in processing their feedback for at least an hour. Thus, it is natural that, even if unintentionally, learners remembered input they had just been processing.

Descriptive statistics also showed this immediate accuracy improvement was partially retained in T3 (new test or delayed post-test) in the two groups receiving external explicit feedback (DIR and REF groups). From test 2 to test 3, the DIR and REF groups increased the error mean that had decreased for both groups from test 1 to test 2. The increased error means of both groups in test 3, however, were still below (DIR-G) or equal to (REF-G) their error mean in test 1. This may suggest that some of the accuracy improvement in test 2 was retained until test 3. This finding (from descriptive statistics only) also supports previous studies' (Bitchener, 2008; Bitchener & Knoch, 2008, 2009a, 2009b; Sheen, 2007; Storch, 2009) findings of accuracy improvements being retained in new writing or a delayed post-test. Contrariwise, the numbers of errors in SELF-G and CONTROL-G from test 2 to test 3 decreased. The behaviour of these two groups will be discussed below, under results with statistical significance.

Although descriptive statistics results showed a positive impact from all noticing-supported ECCs on overall accuracy improvement in the rewriting and to a lesser extent the new writing stages, statistical significance did not support these results. Statistical significance across the three tests was exclusive to: (a) overall and non-grammatical accuracy in DIR-G and (b) non-grammatical accuracy in SELF-G. Concerning the selected seven error types, a series of ANOVAs revealed that except for spelling in the DIR and CONTROL groups, the error type did not influence accuracy performance. The effects of noticing-supported ECCs on accuracy improvement supported by inferential statistics are summarized and further discussed in the following findings.

Effects of Noticing-Supported DIR EC on Overall and N-GR Accuracy

Statistical significance across the three tests was found in DIR-G for overall and non-grammatical accuracy, the latter influencing the former. This finding agrees with the conclusions of previous studies regarding the efficacy of DIR-WCF in post and delayed-post writing tasks (Chandler, 2003), on the advantages of DIR over REF (Sachs & Polio, 2007; Santos et al., 2010) ECs. The reasons for the success of DIR-ECC for non-grammatical accuracy may be various: (a) corrections are more explicit in DIR-ECC than in REF and SELF-ECCs. Manchón (2011b) and Sheen (2010) regard the degree of explicitness of WCF as one of the most influential factors for its success; (b) locating errors and their corresponding corrections is easier and less confusing in DIR than in REF and SELF-ECCs. Santos et al. (2010) explain that corrections in DIR-ECC are more salient than in any other type of feedback; (c) the number of errors or changes in DIR-ECC is also less than in REF-ECC; (d) learners' familiarity with this technique is likely to have an effect too; (e) DIR-ECC meets learners' expectations of clear and direct correction of every single error; (f) computer-mediated correction might have overcome the messy corrections of DIR-EC using pen and paper.

Effects of SELF-EC on N-GR Accuracy

The fact that SELF-G (with only noticing opportunities and self-provided or self-initiated feedback) was the only experimental group that improved its non-grammatical accuracy across the three essays suggests that accuracy improvement might not be a consequence of ECCs but of noticing opportunities. Although supported by inferential statistics (statistical significance was found for non-grammatical accuracy improvement in SELF-G between T1 and T3), this conclusion lacks strength due to the absence of an additional experimental group (one receiving only ECC) for each of the three tested ECCs in the design. Accounting for such conditions however, would have made the study difficult to conduct. Self-correction is a frequently recommended correction technique in the SLA literature and language teaching practice. However, evidence for its efficacy is rarely provided. The significant statistical support for non-grammatical improvement in SELF-ECC refutes Van Beuningen et al.'s (2012) conclusion that WCF is more beneficial for learning than self-correction and sheer writing practice.

Lack of Statistically Significant Results for Grammatical Features

This finding might lend support to Truscott's claim for the inefficacy of WCF for grammatical features in the three tested ECCs. Statistical significance being found exclusively for DIR-G (in overall and non-grammatical accuracy) and SELF-G (in non-grammatical accuracy) means little if any effect of noticing-supported ECCs on grammar accuracy improvement. There were no effects in terms of grammatical accuracy improvement, even in descriptive statistics. This outcome contradicts Van Beuningen et al.´s (2012) conclusion about the efficacy of comprehensive WCF (specifically direct comprehensive WCF) for-improvement of grammatical errors.

Support for Truscott's claims in this study may also derive from the uncontrolled open writing task, and the comprehensive error correction learners received.

The latter refers to the fact that learners received comprehensive EC of their essays; semi-comprehensive EC was used only for essay accuracy analyses. Among the five studies enquiring into comprehensive WCF, results of this study support Truscott and Hsu's (2008) conclusion regarding the lack of effects of comprehensive WCF on grammatical accuracy. The error categories targeted in Truscott and Hsu were mainly grammatical (orthographical and lexical errors were not corrected) because Truscott's claim opposing the efficacy of WCF refers exclusively to grammatical features. Improvement in only non-grammatical accuracy also supports Van Patten's (1994) claim that, in language processing, content or meaning rather than linguistic features or form is processed first, accuracy is secondary for comprehension, and for WCF at the composing stage too, it could be said.

Excluding Spelling, No Error Type Effects in Writing Accuracy

Regarding the seven selected error types, statistical significance across the three tests was found for spelling in the DIR and CONTROL groups. These results may suggest two things. First, except for spelling in DIR and CONTROL groups, error type in writing accuracy was not influenced by the tested ECCs. Second, learners were able to correct their spelling errors with (DIR-G) and without (CONTROL-G) feedback. The results support Truscott's (2007) statement that "spelling errors are among the most correctable error types because they are relatively simple and can be treated as discrete items" (p. 258). In his 12-error category study, Lalande (1982) also found an 83 per cent improvement for orthographic errors. Spelling accuracy has seldom been explored in WCF research. The high frequency of spelling errors that emerged in learners' essays in this study, however, deserves attention. In the researcher's view, spelling accuracy has been undervalued in L2 writing, despite being a feature that contributes greatly to L2 writing accuracy. The results in the present study suggest not only that spelling is the most frequent error in learners' writing but also one most amenable to correction (Truscott, 2007; Lalande, 1982). In the search for writing accuracy, all features (grammatical and non-grammatical) should be addressed.

Effects of Sheer Practice on N-GR Accuracy

The spelling accuracy improvement found in DIR-G (with external explicit feedback) and CONTROL-G (without any type of feedback) is interesting. Spelling improvement in DIR-G vs. REF-G (both receiving external explicit feedback) could be explained by the richness of input in the REF condition that attracted learners' attention to other more relevant features, rather than spelling. However, spelling being significant in CONTROL-G is different. That CONTROL-G managed to improve spelling accuracy without feedback and without noticing opportunities suggests that mere practice might be enough to improve this specific error type. Truscott (1996) claims learners' time and effort would be more productively spent on writing practice. Sheen, Wright, and Moldawa (2009) study did not find evidence for the efficacy of sheer practice. Considering spelling only, this study may contribute some evidence. This finding, also reinforces task proponents' claim for the usefulness of task repetition (Lynch, & Maclean, 2001). Recalling Van Patten's (1994) suggestion that different language aspects require different amounts of

attention, the results suggest that spelling may only require attention at the 'noticing level' to see improvement. If spelling is one of the most treatable errors, teachers might now have justified reasons to pass the responsibility for spelling correction to learners themselves.

Conclusion

The limitations of narrowly focused experimental studies to respond to Truscott's claims was the main concern of this paper. Thus, rather than taking a stand on the grammar correction debate, the researcher aimed to design a more ecological study on the effects of different noticing-supported ECCs on accuracy performance. With this in mind, an effort was made to tackle comprehensive EC, eventually replaced by semi-comprehensive EC. The methodological challenges this EC technique entails made it unfeasible with the open, uncontrolled, learner-centred, 300-word opinion essay used in this study. The combination of various uncontrolled conditions rather than the technique itself might have constrained the feasibility of comprehensive EC in this study. A deeper analysis of the few studies (Ellis et al., 2008; Ferris, 2006; Sheen et al., 2009; Truscott & Hsu, 2008; Van Beuningen et al., 2008, 2012) claiming to work with comprehensive EC, their challenges, strategies, strengths and limitations will be dealt with in a future paper. As far as this study is concerned, the findings may support Truscott's claim for the inefficiency of (semi-comprehensive in this case) WCF for grammatical accuracy. Different from Truscott, the researcher does not claim that WCF is ineffective nor that teachers should stop providing or investigating WCF. The researcher draws attention to (a) a better understanding of Truscott's claims; his arguments are solid and merit closer attention and (b) the fact that narrowly focused experimental studies and Truscott's claims are incompatible. Lee (2011) points out that teachers need to work smarter and not harder in responding to student writing. Traditional forms of WCF may be improved concerning time and energy consumption. A second issue this study draws attention to is that learning does not take place by simply looking at teachers' corrections. Due to learners' limited processing capacity, they need to be trained in attention and noticing to "be selective and … strategically allocated and managed" (Izumi, 2013 p. 35). N-GR accuracy improvement in the SELF and CONTROL ECCs apparently supports the small effect of teacher-provided error correction. Both, the SELF and CONTROL groups received no external explicit EC. Despite this, the learners in those groups improved their non-grammatical accuracy. Truscott (2007) underlines 'error correction' has been used too broadly to include all error types. Error correction may be effective for improving certain errors of a non-grammatical type. Evidence for this claim was found in this study. If non-grammatical errors are more amenable to correction and contribute to writing accuracy, teachers could pass this responsibility over to learners. A better interpretation of the results might be that, after the first composing stage, noticing opportunities should be added to the writing process. Two reasons justify this; first, the above results show that after their first composing, learners are still able to improve their written accuracy by themselves, even if it is only the accuracy of non-grammatical treatable errors. The second reason refers to attention being limited, which obliges learners to distribute their time across task stages; in the composing stage, learners concentrate on conveying meaning, so, another stage is needed to address their attention to form. Results confirmed what

Evans, Hartshorn, McCollum, & Wolfersberger (2010, p. 446) remark on, we do not want to know whether providing WCF is efficient or not, we want to know how we can best help students write more accurately. Researchers might also be approaching what is a practical problem as a theoretical one. Polio (2012) states that error correction is worth investigating "at a practical level even without reference to specific theories" (p. 376), simply because it is a pedagogical practice prevalent in all learning contexts and consumes a lot of time.

Limitations of this study are various: multiple drafting, whole-class feedback and learners with lower language proficiency are still not accounted for in this design; error rate analysis is not the only way to measure accuracy, and; an additional group without noticing opportunities for each of the ECCs included in the design would have been desirable. Directions for future research may include a more systematical approach to comprehensive EC. If it is feasible for research, agreement might first be necessary on the categorisation of different error types. Longitudinal studies on learners' accuracy performance after receiving noticing treatment are also essential. Like Evans et al. (2010) and Lee (2011), the researcher believes that teachers should make EC an essential component of teaching and learning and should continue to enquire into the best ways to help learners improve their writing skill. Evans et al. (2010) state "there are scientific and ethical reasons … to continue research on correction" (447). After all, Johnson (1988) states that "the question of how to provide successful feedback is no less perplexing than the question of how to facilitate successful ... learning" (p. 95).

References

Adams, R. (2003). L2 output, reformulation and noticing: Implications for IL development. *Language Teaching Research, 7*(3), 347–376.

Allwright, R.L., Woodley, M.P., & Allwright, J.M. (1988). Investigating reformulation as a practical strategy for the teaching of academic writing. *Applied Linguistics, 9*(3), 236–256.

Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing, 9*, 227–258.

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*, 102–118.

Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research, 12*, 409–431.

Bitchener, J., & Knoch, U. (2009a). The contribution of written corrective feedback to language development: A ten-month investigation. *Applied Linguistics, 31,* 193–214.

Bitchener, J., & Knoch, U. (2009b). The relative effectiveness of different types of direct written corrective feedback. *System 37,* 322–29.

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing, 14*, 191–205.

Bruton, A. (2009). Improving accuracy is not the only reason for writing, and even if it were… *System, 37,* 600–613.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing, 12*, 267–296.

Cohen, A.D. (1983). *Reformulating second-language compositions: A potential source of input for the learner*. (ERIC ED 228 866).

Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System, 36*, 353–371.

Evans, N. W., Hartshorn, K. J., McCollum, R. M., & Wolfersberger, M. (2010). Contextualizing corrective feedback in second language writing pedagogy. *Language Teaching Research, 14*(4), 445–463.

Fathman, A., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178–190). Cambridge: Cambridge University Press.

Ferris, D. (2004). The 'grammar correction' debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime... ?). *Journal of Second Language Writing, 13*, 49–62.

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81–104). Cambridge: Cambridge University Press.

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes. How explicit does it need to be? *Journal of Second Language Writing, 10*, 161–184.

Godfroid, A., Housen, A., & Boers, F. (2010). A procedure for testing the noticing hypothesis in the context of vocabulary acquisition. In M. Pütz & L. Sicola, (Eds.), *Cognitive processing in second language acquisition. Inside the learner's mind* (pp. 169–198). Amsterdam: John Benjamins Publishing Company.

Guénette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing, 16*, 40–53.

Hanaoka, O. (2007). Output, noticing, and learning: An investigation into the role of spontaneous attention to form in a four-stage writing task. *Language Teaching Research, 11*(4), 459–479.

Hanaoka, O., & Izumi, S. (2012). Noticing and uptake: Addressing pre-articulated covert problems in L2 writing. *Journal of Second Language Writing, 21*, 332–347.

Izumi, S. (2013). Noticing and L2 development: Theoretical, empirical and pedagogical issues. In J.M. Bergsleithner, A.N. Frota & J.K. Yoshioka (Eds.), *Noticing and second language acquisition: studies in honor of Richard Schmidt* (pp. 25–38). Honolulu: University of Hawaii, National Foreign Language Resource Center.

Johnson, K. (1988). Mistake correction. *ELT Journal, 42*(2), 89–96.

Lalande, J. F. (1982). Reducing composition errors: An experiment. *Modern Language Journal, 66*, 140–149.

Lee, I. (2011). Working smarter, not working harder: Revisiting teacher feedback in the L2 writing classroom. *The Canadian Modern Language Review, 67*(3), 377–399.

Leow, R. (1997). Attention, awareness and foreign language behaviour. *Language Learning, 47*(3), 467–505.

Lynch, T., & Maclean, J. (2001). A case of exercising: effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan & M.Swain (Eds.) *Researching pedagogic tasks: Second language learning, teaching and testing*. (pp. 141–162). New York: Routledge.

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics, 27*(3), 405–430.

Manchón, R. M. (2011a). The language learning potential of writing in foreign language contexts: Lessons from research. In T. Cimasko & M. Reichelt (Eds.), *Foreign language instruction. Principles and practices* (pp. 44–64). USA: Parlor Press.

Manchón, R. M. (2011b). Writing-to-learn the language. In R.M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 61–82). Amsterdam: John Benjamins Publishing Company.

Manchón, R.M. (2013). Teaching writing. In C. Chapelle (Ed.), *The encyclopaedia of applied linguistics* (pp. 2−7). USA: Blackwell Publishing.

Park, E.S. (2011). Learner-generated noticing of written L2 input: What do learners notice and why? *Language Learning, 61*(1), 146–186.

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing, 21*, 375–389.

Qi, D. S., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing, 10*(4), 277–303.

Reinders, H. W. (2005). *The effects of different task types on L2 learners' intake and acquisition of two grammatical structures* (Unpublished doctoral thesis). University of Auckland, Auckland.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45*(2), 283–331.

Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on an L2 writing revision task. *Studies in Second Language* Acquisition, *29*, 67–100.

Santos, M., López-Serrano, S., & Manchón, R.M. (2010). The differential effect of two types of direct written corrective feedback on noticing and uptake: Reformulation vs. error correction. *International Journal of English Studies, 10*(1), 131–154.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129–158.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu, Hawai'i: University of Hawai'i, Second Language Teaching and Curriculum Centre.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly, 41*, 255–283.

Sheen, Y. (2010). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition, 32*, 203–234.

Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System, 37*, 556–569.

Solares-Altamirano, M.E. (2016). *The effects of different error correction conditions on learner-initiated noticing* (Unpublished PhD thesis). Lancaster University, United Kingdom.

Storch, N. (2009). *Comparing type of feedback and processing mode: Pair versus individual processing of feedback on writing*. Paper presented at the Second Language Writing Symposium, Phoenix, Arizona.

Storch, N. (2010). Critical feedback on written corrective feedback. *International Journal of English Studies, 10*(2), 29–46.

Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing. Case Studies. *Studies in Second Language Acquisition, 32*, 303–334.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step toward second language learning. *Applied Linguistics, 16*, 371–391.

Swain, M., & Lapkin, S. (2007). The distributed nature of second language learning: Neil's perspective. In S. Fotos & H. Nassaji (Eds.), *Form-focused instruction and teacher education. Studies in honour of Rod Ellis* (pp. 73–86). Oxford: Oxford University Press.

Tomlin, R., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16*, 183–203.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*(2), 327–369.

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing, 16*, 255–272.

Truscott, J., & Hsu, A. Y. (2008). Error correction, revision, and learning. *Journal of Second Language Writing, 17*, 292–305.

Van Beuningen, C. (2011). *The effectiveness of comprehensive corrective feedback in second language writing* (Unpublished PhD thesis). University of Amsterdam, Amsterdam.

Van Beuningen, C., De Jong, N., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learners' written accuracy. *ITL International Journal of Applied Linguistics, 156*, 279–296.

Van Beuningen, C., De Jong, N., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning, 62*(1), 1–41.

Van Patten, B. (1994). Evaluating the role of consciousness in second language acquisition: Terms, linguistic features & research methodology. *AILA Review,11*, 27–36.

Williams, J. (2001). Learner-generated attention to form. In R. Ellis (Ed.), *Form focused instruction and second language learning* (pp. 303–346). Malden, MA: Blackwell.

Yang, L., & Zhang, L. (2010). Exploring the role of reformulations and a model text in EFL students' writing performance. *Language Teaching Research, 14*(4), 464-484.

Fark Etme Destekli Hata Düzeltmenin Öğrencinin Doğru Yazması Üzerine Etkileri

Özet

Neredeyse yirmi beş yıldır, araştırmacılar Truscott'un yazılı düzeltici geribildirimin (YDG) dilbilgisi gelişimi için yetersiz olduğu iddiasını çürütmeye çalışmışlardır. Çok sayıda çabaya rağmen, YDG'nin dilbilgisi öğreniminde oynadığı rol halen çözülememiştir. Truscott'un iddiasını çürütmeye çalışan ilk çalışmalar, son çalışmaların üstesinden geldiği yöntemsel ve etik kusurlar içeriyordu. Yine de Truscott'un iddiası tamamen çürütülememiştir. Bu çalışma, bunun olası nedenlerini tartışmakta ve fark etmeyi destekleyen hata düzeltme koşullarının (HDK) yazma doğruluğu üzerindeki etkilerini inceleyen yarı deneysel bir araştırmayla desteklemektedir. Araştırma açık, kontrolsüz ve öğrenci merkezli bir yazma göreviyle YDG'yi kapsamlı bir şekilde ele almaktadır. Geribildirimin her aşamasında öğrenci tarafından başlatılan fark etme olanakları HDK'lara eklenerek yazma doğruluğunun etkilenip etkilenmediğini araştırılmıştır. YDG üzerine yapılan çok sayıda araştırmadan pek azı geribildirim esnasında fark etme konusunu incelemiştir Çalışmanın kuramsal çerçevesini "fark etme" "yazmanın dil öğrenmede potansiyeli", "öğrenmek için yazma" ve "edinim için geribildirim" kavramları oluşturmaktadır. Sonuçlar: Truscott'un YDG'nin dilbilgisel gelişimde yetersizliği iddiasına destek verebilecek istatistiksel olarak anlamlı olmayan bulgular ve KONTROL ve ÖZ-HDK'lar arasındaki istatistiksel olarak anlamlı bulgular olmak üzere iki temel başlık altında tartışılmıştır. Bu bulgulardan ilki, Truscott'un öğrencilerin zaman ve çabasının yazma pratiğine odaklanarak daha verimli bir şekilde değerlendirilebileceği iddiasını destekleyebilir. İkincisi ise, yazıda doğru kullanımı geliştirmenin YDG tekniğine bağlı olmadığını, ancak fark etme fırsatlarına bağlı olabileceğini düşündürmektedir.

Anahtar Kelimeler: Yazılı düzeltici geribildirim, hata düzeltme, kapsamlı/odaklanmamış hata düzeltme, öğrenen tarafından başlatılan fark etme