Language Learning & Technology
ISSN 1094-3501
June 2021, Volume 25, Issue 2
pp. 135–157

ARTICLE

Can a 'pedagogical' spellchecker improve spelling accuracy in L2 Spanish?

Miguel Blázquez-Carretero, University of the Philippines Diliman
Robert Woore, University of Oxford

Abstract

Accurate spelling matters for L2 learners: It facilitates communication, affects other aspects of the writing process, and is an important assessment criterion. However, even in phonologically transparent writing systems like Spanish, L2 learners experience spelling difficulties. Nonetheless, explicit spelling instruction appears to be neglected by L2 teachers. Synchronous written corrective feedback, provided automatically by computerised spellcheckers, is one way of providing such instruction without cost to teaching time. However, evidence concerning the effectiveness of such feedback is mixed. Further, existing spellcheckers, designed for L1 speakers, present various problems for L2 learners. The current study reports on an experimental trial of a Pedagogic Spellchecker (PSC), developed specifically for L2 learners. In all, 107 adult learners of Spanish as a Foreign Language were block randomised into three treatment groups. All groups completed a short transcription task on five consecutive days. One group received feedback from the PSC; another received feedback from Microsoft Word spellchecker; the third received no feedback. Pre- and post-test data showed that the PSC group progressed significantly more in spelling accuracy than the other groups, with a large effect size. Nonetheless, Microsoft Word spellchecker reduced errors on spelling forms that it did not autocorrect. Pedagogical and theoretical implications are discussed.

Keywords: *Spellchecking Software; Written Corrective Feedback (WCF); Spanish as a Foreign Language (SFL); Computer-Assisted Language Learning (CALL)*

Language(s) Learned in This Study: *Spanish*

APA Citation: Blázquez-Carretero, M., & Woore, R. (2021). Can a 'pedagogical' spellchecker improve spelling accuracy in L2 Spanish? *Language Learning & Technology*, *25*(2), 135–157. http://hdl.handle.net/10125/73437

Introduction

Accurate spelling is important. First, it affects communication: Misspelled words may make a text harder to read and distract readers from the intended message (Graham et al., 2008). Second, spelling is interlinked with other aspects of producing written texts: Given limited working memory capacity, focussing attention on how to spell a word may interfere with conceptualising ideas and formulating the language needed to express them. Further, as Stirling (2011) suggests, students may avoid using words which they are unsure how to spell, thereby limiting their lexical range or changing their intended message. Finally, accurate spelling matters because it is a criterion by which learners' written work is evaluated in high-stakes examinations (Rodríguez et al., 1997).

Nonetheless, spelling often appears to be neglected in second language (L2) teaching (Berkel, 2004). Perhaps, Berkel speculates, this is because "it is assumed that everyone knows how to spell ... and L2 pupils are supposed to learn without instruction" (p. 239)—in other words, they are expected to learn incidentally. In the specific case of teaching Spanish as a Foreign Language (SFL), too, there is evidence that spelling instruction is neglected. For example, based on anecdotal evidence from several SFL teaching environments, Sánchez Jiménez (2013) found that teachers tended to prioritize other aspects of language

over spelling. This may reflect teaching approaches which prioritize communication of meaning over formal accuracy (Burston, 2001).

Spelling in L2 Spanish

Spanish is amongst the 'shallowest' (most phonologically transparent) European writing systems (Seymour et al., 2003). Therefore, some SFL teachers may expect the spelling system to be easy to acquire incidentally, compared to 'deeper' writing systems such as English. Learners of Spanish have fewer individual symbol-sound correspondences to learn; having learnt them, they can (in theory) spell words that they know orally, but have not yet encountered in writing, with high accuracy (Barry, 1992).

However, L2 Spanish learners do make frequent spelling errors. For example, Sánchez Jiménez (2010) analysed 258 compositions written by Filipino SFL learners. Out of 30,194 words written, 10.5% were misspelled. Similarly, Blazquez and Fan (2019) analysed 30 compositions from the CORANE corpus (Cestero et al., 2009), written by SFL students from various first language (L1) backgrounds. Out of 4,458 words, 8.6% were misspelled.

Some challenges of Spanish spelling derive from the language's orthography itself. Despite its shallowness, it does have complexities and inconsistencies, posing challenges for L2 learners irrespective of L1 background. Examples include graphemes with multiple pronunciations (e.g., <c> in *cinto*, /θinto/, 'belt' versus *canto*, /kanto/, 'song') and heterographic homophones (e.g., *más*, /mas/, 'more' versus *mas*, /mas/, 'but'). However, for learners already literate in their L1, the difficulty of mastering an L2 writing system derives not only from that system's internal features, but also from its relationship to the L1. This may explain Sánchez Jiménez's (2010) finding that spelling errors made by SFL learners differ from those made by native Spanish speakers. For example, English-speaking learners may spell the Spanish word *asociación* with a double <ss> (*<associación>) by analogy to their L1 (*association*). Such problems could be characterised as *interference*, negative transfer from their L1 (Odlin, 1989).

Written Corrective Feedback and L2 Spelling

Where teachers do explicitly focus on spelling, one common means of doing so is via written corrective feedback (WCF) on pupils' written work. WCF has been the focus of much research attention and controversy (see Bitchener & Ferris, 2012 for a recent overview). However, Kang and Han's (2015) recent meta-analysis of experimental studies concluded that WCF—provided by a teacher—"does have a substantive effect on L2 written accuracy" (p. 10) in immediate post-tests, with a small to moderate effect size. There was also some evidence that the accuracy improvements persisted in delayed post-tests. However, these findings were based specifically on grammatical accuracy, not spelling accuracy more broadly.

Within the field of Computer-Assisted Language Learning (CALL), a considerable body of research has also investigated the effects of computer-mediated WCF, as extensively reviewed by Heift (2017) and Heift and Hegelheimer (2017). Again, most research in this area has focussed on grammar development, but there have been some studies of L2 spelling, as we will explore below.

One complicating factor for research on WCF is that feedback can take various forms, some possibly more effective than others. WCF has generally been categorized using four main dimensions. The first concerns whether the feedback is focused (addressing specific language features, such as perfect tense forms) or 'unfocused' (addressing all errors). Summarising research to date, Shintani and Aubrey (2016) note that most studies investigating focused WCF have found positive effects on written accuracy in both immediate and delayed post-tests; studies on unfocused WCF are less frequent and have produced more mixed findings.

Second, WCF can be direct (providing corrections) or indirect (identifying errors but not correcting them). The latter may be thought more likely to encourage deep processing (Craik & Lockhart, 1972), as learners draw on existing knowledge and/or external resources to correct errors for themselves; in turn, this deep

processing may promote learning. Conversely, learners who are given indirect feedback may fail to solve problems correctly, substituting one error for another or getting stuck entirely. Direct feedback, by contrast, ensures that learners have a definitive, correct version of targeted items for future reference and learning. Kang and Han's (2015) meta-analysis of teacher-provided feedback found a positive effect of both direct and indirect WCF on writing accuracy. No significant difference was found between the two types of feedback; however, the overall effect size was greater for direct WCF (g = 0.60) than indirect (g = 0.36).

Third, WCF can be either synchronous (provided while the learner is still writing the text) or asynchronous (provided after text completion). The advent of online multi-author text editing software, such as *Google Docs*, has facilitated the provision of synchronous WCF by teachers because they are able to watch learners write and to give feedback on their work in real time. In the only experimental study that we know of comparing these two modes of feedback, Shintani and Aubrey (2016) found that both asynchronous and synchronous WCF led to greater accuracy gains (on a targeted grammatical structure) than a control (no feedback) condition. However, synchronous WCF had "greater and more durable effects" than asynchronous feedback (p. 309). The authors interpret these findings within a sociocultural framework, hypothesising that synchronous WCF allows teachers to optimize the scaffolding provided to learners at each stage in the writing process.

Fourth, corrective feedback may be supplied with or without metalinguistic information which explains to learners the nature of their errors. In relation to spelling, this explanation might comprise a brief statement of any relevant rules (e.g., 'the sound /s/ is represented in Spanish only by single <s>, never double <ss>'). In theoretical terms, the assumption is that such information may contribute to learners' explicit, declarative knowledge, which can subsequently—through practice—be proceduralized and ultimately automatized within a skill acquisition framework (DeKeyser, 2015). Several CALL studies have compared metalinguistic feedback to other forms of WCF (e.g., Heift, 2004; Lado et al., 2014; Nagata, 1993; Rosa & Leow, 2004). Reviewing this body of research, Heift and Hegelheimer (2017) conclude that overall, metalinguistic feedback appears to be beneficial for L2 learning, at least in the immediate term; however, they also assert that further research is needed, particularly in relation to longer-term effects (as measured on delayed post-tests), where current evidence is less conclusive.

Spellchecking Software

In Shintani and Aubrey's (2016) study, synchronous WCF was provided by a human teacher to a group of 25 students. However, the writing tasks were highly controlled, with feedback focussing solely on a particular grammatical structure. It is unclear how feasibly teachers could provide real-time WCF on more open-ended tasks with less predictable errors. Moreover, students sometimes complete writing tasks outside class (e.g., for homework) and therefore at differing times, making it difficult for teachers to provide real-time feedback.

One solution to such problems could be provided by the spellchecking functions built into many word processing packages. As noted by Heift (2017), such generic spellcheckers are widely used by L2 learners and "[w]ith little controversy … are praised for their effectiveness in treating spelling mistakes" (p. 292). However, research into the effects of generic spellcheckers on L2 learning (as reviewed by Heift, 2017) remains at an early stage and there are grounds for caution.

One problem is that such packages, being designed for L1 writers, may be less effective in dealing with errors made by L2 learners (Burston, 1998). For proficient L1 writers, misspellings are often mere slips of the keyboard, or what Corder **(**1967) termed performance errors. By contrast, for L2 learners, misspellings may reflect genuine lack of knowledge of the correct orthographic form—in Corder's terms, competence errors. Indeed, Rimrott and Heift's (2005) analysis of a corpus of writing in L2 German found that 80% of the errors could be classified as competence errors. Furthermore, when the authors submitted the errors to MS Word's generic spellchecker, the competence errors were much less successfully dealt with than the performance errors. Overall, only 52% of errors were successfully corrected, thus questioning the

usefulness of generic spellcheckers for L2 learners.

In one of the few experimental studies to date examining the effects of a generic spellchecker on L2 outcomes, Lin et al. (2017) investigated whether feedback from a custom-made spellchecker would help learners to correct errors and improve their spelling knowledge. A total of 88 Taiwanese university students of L2 English were randomly allocated to four conditions. All participants proofread an English text containing 23 common learner misspellings, and were asked to identify and correct as many of these errors as possible. In one condition ('red-underline'), errors were highlighted by red underlining. In another ('dictionary'), they were red-underlined and a computerised dictionary was available. In a third ('spellcheck') condition, errors were red-underlined and a drop-down menu of possible spellings was provided. Finally, a fourth (control) condition had no spelling aids. Participants' ability to detect and correct spelling errors in similar passages was measured in pre-, post- and delayed post-tests. The groups performed similarly at pre-test. At post-test, however, all three red-underlining groups outperformed the control group in both detecting and correcting errors. In the delayed post-test, all three red-underlining groups again outperformed the control group in *detecting* errors, but the 'dictionary' and 'spellcheck' groups additionally outperformed the control group in *correcting* errors. The authors conclude that generic computerized spellcheckers, besides being convenient aids for L2 writing, can improve learners' spelling knowledge through incidental learning. They concede, however, that the delayed post-test occurred only one day after the treatment, limiting conclusions about the durability of the learning. Further, participants proofread an existing text; outcomes may have differed had learners located and corrected errors in their own L2 writing.

A further limitation of Lin et al.'s (2017) study derives from its exclusive use of a custom-made spellchecker. This differs from the widely-used MS Word spellchecker in an important respect: It lacks an autocorrect function, whereby some errors are corrected automatically, instantaneously and without notification. In the case of performance errors (e.g., if a writer accidentally types **asociación* for *asociación*, even though they actually know the correct form), this functionality may well be useful and time-saving. However, in the case of competence errors, autocorrection may be unhelpful. Swain (1998) argued that 'noticing the gap' between one's own linguistic output and the correct form promotes learning. Autocorrection seems likely to inhibit learners' noticing of their errors (Schmidt, 1990) and thus their ability to learn from them (Lawley, 2015).

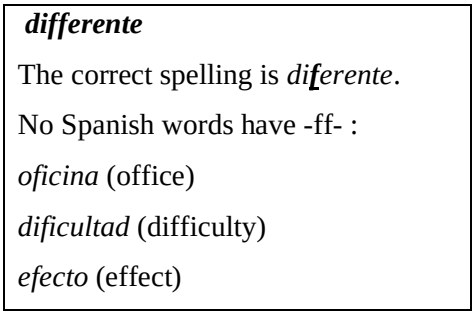
A Pedagogic Spellchecker

Responding to the limitations of generic spellcheckers, there have been calls for corrective software better attuned to the needs of L2 learners, for example by drawing on a database of common learner errors (e.g., Rimrott & Heift, 2005). An early example was the grammar checker for learners of French (also incorporating a spellchecker) developed and evaluated—with positive results—by Burston (2001).

Building on work such as Burtson's, the *Universidad Nacional de Educación a Distancia* (UNED) has developed a pedagogic spellchecker (PSC), designed for teaching spelling to SFL learners. According to the classifications above, it provides direct, unfocussed, synchronous, metalinguistic feedback. It identifies and provides correct forms for spelling errors attested frequently in the CORANE corpus **(**Cestero Mancera & Penadés Martínez, 2009). Additionally, it provides brief, error-specific feedback, explaining relevant spelling rules and giving examples of other words which follow the same pattern (Lawley, 2016). For instance, *diferente* is frequently misspelled by SFL learners as **differente*, presumably reflecting L1 interference (cf. English *different*, French *différent*). Whereas Microsoft Word simply autocorrects this error to *diferente*, the PSC elaborates, as shown in Figure 1 (the PSC's language is Spanish, but an English translation is given here for convenience). The intention is that learners will not only notice the gap between their own output and the correct form, but also develop explicit knowledge about relevant spelling patterns, which can be deployed in future writing.

Figure 1

*PSC Response to *Differente*



Study Aims

The current study explores the effects of three different types of synchronous error correction on the development of L2 Spanish spelling accuracy. Specifically, it compares the outcomes of learners in three groups. One group (MSW) wrote Spanish text using Microsoft Word and its inbuilt spellchecker. The second (PSC) wrote using UNED's word processor and Pedagogic Spellchecker, designed for SFL learners. The third (Control) group used Microsoft Notepad, with no inbuilt spellchecker. The outcome measure was spelling accuracy in new pieces of writing, produced without access to the spellcheckers. The study therefore assessed improvements in participants' spelling knowledge in long-term memory.

Additionally, the study explores the possibility that the two types of error correction provided by MSW Spellchecker (autocorrection versus non-autocorrection) could have differential effects. This follows the hypothesis that autocorrection might hinder participants' noticing of errors (Schmidt, 1990) and thus reduce learning. Conversely, choosing the desired alternative from a drop-down list (as in the case of non-autocorrected errors) may promote noticing and learning. Finally, the study measured the time taken by participants to write using the three word-processing packages, allowing any benefits of the spellcheckers to be weighed against their time costs.

The research questions were:

1. To what extent does synchronous, unfocussed, direct feedback from (a) Microsoft Word (MSW) spellchecker and (b) the Pedagogic Spellchecker (PSC) lead to improvements in spelling accuracy in new writing, as compared to a control group receiving no WCF?
2. To what extent does one of the two spellcheckers (MSW or PSC) lead to greater improvements in spelling accuracy than the other?
3. In MSW, to what extent is there a difference in the effects on spelling accuracy between (a) errors which are autocorrected and (b) those which are not?
4. To what extent is there a difference in the time taken to write L2 text by learners using (a) the MSW spellchecker, (b) the PSC and (c) no spellchecker?

Methods

The study employed an experimental, pre-, post-, delayed post-test design. After pre-testing, participants were allocated in equal proportion to the three groups (MSW, PSC, Control) using block randomisation to ensure an equal balance of Spanish proficiency levels (measured by a placement test) and gender.

Participants

Participation in the current study was voluntary; no reward was given. The initial sample comprised 107 students (72 female, 35 male; mean age 20 years) from 8 intact classes completing SFL Summer courses at University of Castilla-La Mancha. Students on these courses spent one month in Spain, receiving four hours of Spanish lessons daily, five days a week. Teachers reported that, during the current study, around three hours per week were spent on writing in Spanish, but spelling was not taught explicitly. Further, individual WCF was rarely provided, due to high student numbers.

Due to considerable attrition at delayed post-test (with the loss of almost half the sample), and evidence that this attrition was biased (with the loss of weaker spellers from the Control group in particular; see Appendix A), our analysis focused on the pre-/post-test data only. Thus, we were only able to investigate the immediate effects of our interventions, not any longer-term effects.

Participants were of mixed nationalities, L1 backgrounds and proficiency levels in L2 Spanish (ranging from absolute beginner to upper intermediate). The table in Appendix B summarizes the sample characteristics at pre- and post-test.

Instruments

To measure Spanish proficiency, University of Oxford Language Centre's online placement test (http://www.lang.ox.ac.uk/tests/tst_placement_spanish.html) was used. This placement test (PT) is comprised of 50 sentences, each with a missing verb. Participants must select the correct conjugated form from four options. Whilst no published data is available concerning this test's validity or reliability, it has been used successfully by the Language Centre for many years as "a rough, low-stakes guide for placement" of students onto courses of the appropriate level (R. Vanderplank, previously Director of Oxford University Language Centre, personal communication, November 23, 2018).

To measure spelling accuracy, three transcription tasks were used (one version each at pre-, post- and delayed post-test). In each task, participants heard 17 Spanish sentences (Appendix C), audio-recorded by the first author, a Spanish native speaker. Sentences ranged from 4 to 11 words in length (mean length: 6.2 words). Participants typed out the sentences using the basic text editor, Microsoft Notepad. A pilot study suggested that the three tests were of equal difficulty: We found no significant difference between the scores of nine SFL postgraduate students at the University of Oxford on the three transcription tasks ($F(2,8) = 0.017$; $p = .983$; $d = 0.002$).

Our rationale for selecting a transcription task to measure spelling accuracy—rather than other tasks which have been used in CALL research, such as asking participants to write their own text—was because the PSC, still in prototype form, provided feedback only on certain Spanish spelling patterns. Therefore, it was necessary to ensure that participants produced words containing these patterns. We considered using a translation task to constrain their written output (e.g., Rimrott & Heift, 2005), but piloting showed that participants often used avoidance strategies (e.g., synonyms or paraphrase), thus failing to produce the intended forms. We could also have asked participants to spellcheck an existing document (e.g., Lin et al., 2017) containing the targeted spelling patterns. However, this was considered less ecologically valid for assessing spelling accuracy in learners' written language production.

We acknowledge that using a transcription task introduces additional, possibly confounding variables. For example, it places demands on working memory: As argued by Harrison et al. (2016), participants must retain in memory the sentence they have heard, while retrieving the spellings of the component words. To mitigate this problem, following Conroy (2018), transcription items were short in length (and thus unlikely to overburden participants' working memory). Furthermore, due to the random allocation of participants to conditions, any effects of working memory on participant outcomes should apply equally to all three experimental groups, and thus be controlled for. The same is true of other possible confounds associated with the transcription task, such as phonological segmentation skills and keyboard fluency.

Each of the 3 sets of 17 sentences in the transcription tasks contained 26 target spelling errors (TSEs). The criteria for selecting the TSEs were twofold. First, they are frequently misspelled by SFL learners (Blazquez & Fan, 2019; Justicia et al., 1999; Peñalver Castillo, 2009). Second, they follow a spelling convention which is easily summarised as a simple pattern or rule of thumb. The TSEs were embedded in words carefully selected such that half were autocorrected by the MSW spellchecker; for the other half, MSW spellchecker provided a drop-down list of options. The TSEs and words containing them appear in Appendix D.

Short questionnaires elicited participants' L1 and, at each round of testing, the extent to which they had practised writing at home (see Table 2).

Treatment

The treatment consisted of five transcription tasks, identical in format and length to those completed in the pre-, post- and delayed post-tests, but with different sentences. Participants completed these using different text editing software according to their group, as described above: MSW, PSC or Control. The MSW group had their TSE errors either automatically corrected (50% of TSEs) or highlighted, with alternative spellings being suggested via right-click (the other 50%). The PSC group had their errors highlighted and were given brief explanations of each error (Figure 1). The Control group received no error correction or feedback.

Procedure

Data collection took place in Summer 2016. On day 1 (a Friday), participants completed the first questionnaire, online placement test and transcription pre-test as a whole group in a computer lab. No time limit was set for the pre-test (or any of the subsequent tests). This was to reduce pressure or anxiety **(**Liu & Brown, 2015), and to allow explicit knowledge to be drawn upon when writing and checking spellings.

The following Monday was the first day of the treatment, which spanned five consecutive days. Participants received an initial briefing in how to use the respective word processing packages, then completed one transcription task daily. No time limit was set for the transcriptions, but participants were asked to time their own performance using an online stopwatch (http://cronometro-online.chronme.com). Whilst this self-timing method has potential limitations in terms of reliability, to mitigate this risk, clear instructions were given and the self-timing was overseen by the first author. Further, as each transcription took around ten minutes to complete, even several seconds of measurement error would be small in relation to the overall timing.

The post-test transcription task (and brief questionnaire) were administered on the Monday following the end of the treatment. Finally, the delayed post-test and further brief questionnaire were administered one week later.

Data analysis

When marking the transcription tasks, a spelling error was defined as any sequence of letters between spaces that does not appear in the dictionary of the *Real Academia Española* **(**RAE, 2014**)**. Errors were classified as either TSEs (if they were one of the 26 targeted errors) or 'Other Spelling Errors' (OSEs). The inclusion of OSEs reflects the possibility that participants might improve accuracy on wider spelling patterns beyond those targeted by the PSC. When participants omitted a word containing a TSE, or wrote a different word, this was classified as Not Written (NW). In total (across the three tests), only 6% of responses were classified as NW, equally distributed across the groups. We have excluded these responses from the analyses.

Marking was completed by the first author. To check reliability, a second rater re-marked 12 randomly selected sets of pre-, post- and delayed post-test transcription tasks (5% of the 245 sets of transcriptions completed in total). There was total agreement between the two raters.

Results

Preliminaries

There was some attrition of our sample between pre- and post-test: Of the 107 initial participants, 82 (76.6%) completed both rounds of data collection. Table 1 indicates the numbers, in each group, of (a) pre-test participants; (b) post-test participants, defined as those who completed both pre- and post-tests; (c) number of attritors at post-test; and (d) attrition rate at post-test. It also shows the pre-test spelling accuracy (number of misspellings in TSE items) of the attritors and non-attritors, both overall and in each individual group (a lower number indicates fewer errors and therefore greater accuracy).

Table 1

Numbers and Pre-Test Scores of Attritors and Non-Attritors at Post-Test

					Pre-test scores of attritors		Pre-test scores of non-attritors	
Group	***N* (pre)**	***N* (post)**	***N* Attrition**	**% Attrition**	***M***	***SD***	***M***	***SD***
PSC	35	27	8	22.9	10.3	7.1	11.0	6.5
MW	36	28	8	22.2	9.0	4.4	11.7	5.4
Control	36	27	9	25.0	10.7	5.4	10.5	5.4
Total	107	82	25	23.4	10.0	5.5	11.1	5.7

Given the loss of almost one quarter of our sample at post-test, there was a risk of attrition bias. An independent samples t-test was therefore conducted to compare the pre-test scores of the attritors and non-attritors, in the sample as a whole and in each group. No significant difference was found either for the whole sample, $t(105) = 0.827$, $p = .410$, or in any of the individual groups, PSC: $t(33) = 0.267$, $p = .791$; MSW: $t(34) = 1.307$, $p = .200$; Control: $t(34) = 0.071$, $p = .944$. Thus, no evidence of attrition bias was found.

We next checked participants' characteristics in each group, in terms of gender, L1 background, and L2 Spanish proficiency as measured by placement test (PT) scores (Table 2). We simplified L1 background into a binary variable (English or 'other'), given that almost all participants had either English or Chinese L1, with only very few ($n = 2$) French speakers. Additionally, we checked how many participants in each group self-reported (in the post-test questionnaire) that they had practised spelling at home, outside the timetabled treatment sessions, since this might have affected outcomes. Again, we reduced this to a binary variable (those who had practised at home versus those who had not), although we acknowledge the limitations of simplifying the data in this way.

Table 2

Characteristics of the Pre-/Post-Test Sample

		PT score (out of 50)		L1 background		Gender		Practiced at home	
Group	***N***	***M***	***SD***	**English**	**Other**	**F**	**M**	**Yes**	**No**
PSC	27	30.0	9.1	26	1	21	6	11	16
MW	28	31.5	7.5	21	7	19	9	10	18
Control	27	30.1	9.0	22	5	16	11	10	17
Total	82	30.6	8.5	69	13	56	26	31	51

Chi square tests revealed no significant relationship between group membership and either: (a) gender, $\chi^2(2)$

= 2.142, $p = .343$; (b) L1 background, $\chi^2(2) = 4.887$, $p = .087$; or (c) whether or not participants had practised at home, $\chi^2(2) = 0.158$, $p = .924$. Further, a one-way ANOVA found no significant difference between the groups' L2 Spanish proficiency scores, $F(2, 79) = 0.279$, $p = .758$. Thus, we found no evidence that our experimental groups differed significantly at baseline on any observable characteristics.

TSE Scores

We now present the analyses of participants' spelling accuracy on the transcription tests. Table 3 and Figure 2 show the mean number of TSE errors (out of 26) made by each group at pre- and post-test (with lower values indicating higher accuracy).

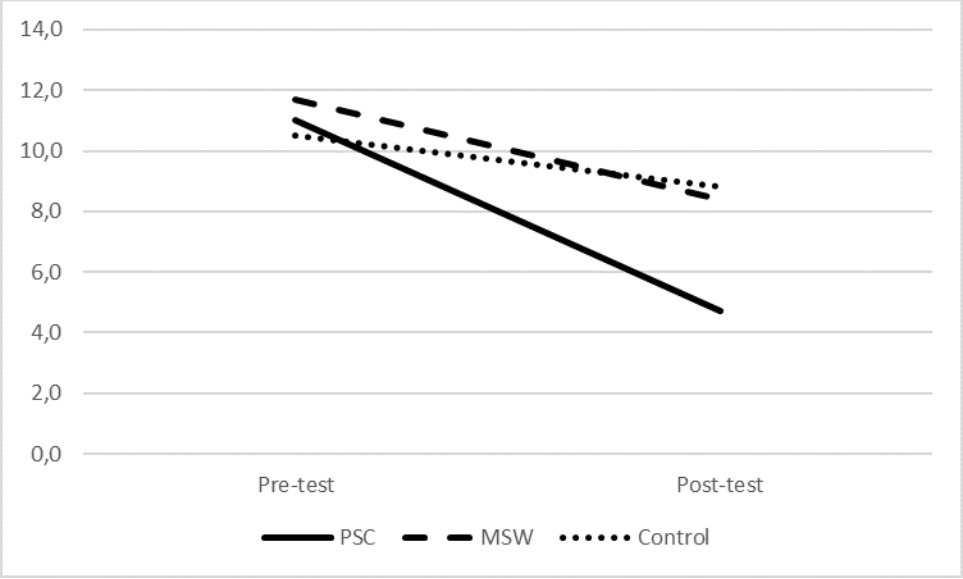
Table 3

Mean Numbers of TSE and OSE at Pre- and Post-Test

Group (*n*)	TSE						OSE					
	PSC (27)		MSW (28)		Control (27)		PSC (26)		MSW (28)		Control (27)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pre-test	11.0	6.5	11.7	5.4	10.5	5.4	12.8	6.7	13.0	6.0	12.2	6.0
Post-test	4.7	4.2	8.4	6.5	8.8	6.1	10.4	6.9	9.8	5.3	11.8	6.8

Figure 2

Mean Numbers of TSE at Pre- and Post-Test



Paired samples t-tests found that TSE scores decreased significantly (i.e. accuracy on targeted spelling patterns increased) for both the PSC and MSW groups, $t(26) = 6.043$, $p < .001$, $d = 1.16$ and $t(27) = 3.589$; $p = .001$, $d = 0.68$ respectively. For the Control group, the decrease in TSE scores approached significance, $t(26) = 2.044$, $p = .051$, $d = 0.39$.

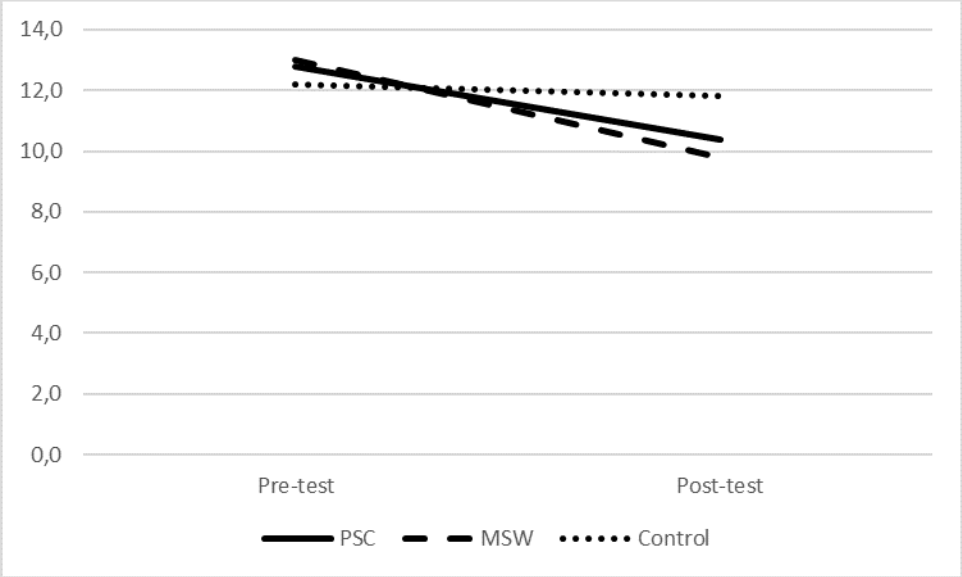
We next compared the groups' performance at each time point. At pre-test, a one-way ANOVA found no significant difference between the groups' TSE scores, $F(2, 79) = 0.302$, $p = .740$. We then conducted a one-way ANCOVA to determine whether there was any significant difference between the three groups' TSE scores at post-test, after controlling for their pre-test TSE scores (entered as a covariate). There was a significant effect of group with a large effect size, $F(2, 78) = 7.083$, $p = .001$, $\eta_p^2 = 0.154$. Post-hoc tests (Bonferroni pairwise comparisons) showed that, after controlling for pre-test scores, the PSC group scored significantly lower (i.e. was significantly more accurate in targeted spelling patterns) than both the MSW group, with a large effect size ($p = .024$, $d = 0.88$), and the Control group, with a very large effect size ($p = .002$, $d = 1.08$). Comparing the MSW and Control groups, the difference between their post-test scores was small and non-significant ($p = 1.000$, $d = 0.19$). Thus, the PSC appears to have conferred a large and significant advantage over both the other conditions in terms of improving participants' accuracy on targeted spelling patterns.

OSE scores

We repeated the above analyses for the 'other spelling errors' (OSE). Table 3 and Figure 3 show the mean number of OSE made by each group at pre- and post-test (with a lower score again indicating greater accuracy)[1].

Figure 3

Mean Numbers of OSE at Pre- and Post-Test



Paired samples t-tests found that OSE scores decreased significantly (i.e. there was an increase in accuracy on non-targeted spelling patterns) for the MSW group, $t(27) = 2.950$, $p = .006$, $d = 0.56$. For the PSC group, the difference between pre- and post-test scores was on the boundary of significance, $t(25) = 2.062$, $p = .05$, $d = 0.41$. The Control group showed no significant difference between pre- and post-test scores, $t(26) = 0.437$, $p = .666$, $d = 0.09$.

A one-way ANOVA found no significant differences between the groups' pre-test OSE scores, $F(2, 79) = 0.412$, $p = .664$. We then conducted a one-way ANCOVA to test for significant between-group differences in post-test OSE scores, after controlling for pre-test OSE scores. No significant effect of group was found, $F(2, 78) = 1.459$, $p = .239$. In other words, we found no evidence that any of the conditions led to greater accuracy on non-targeted spelling forms than the others.

Effects of Autocorrection

Research question 3 investigated the possibility that, for the MSW group, there may be differential effects on spelling accuracy for (a) those errors that are autocorrected by the spellchecking software (henceforth autocorrected errors) and (b) those that are not, for which a list of alternatives is provided in a drop-down menu. Recall that the TSE in the current study were selected such that half were autocorrected by MSW Spellchecker, and half not.

The comparison between autocorrected and non-autocorrected TSE was made for all three experimental groups (even though there was no autocorrection in the PSC or Control groups). This allowed us to check whether any significant difference found between the two categories of TSE in the MSW group could actually be attributed to the autocorrection, rather than being an artefact of the items themselves.

Table 4 and Figure 4 show the numbers of autocorrected and non-autocorrected TSE (out of 13) made by participants in each group at pre- and post-test. For each individual group, we conducted paired samples t-tests to check for statistically significant differences in the numbers of both autocorrected and non-autocorrected errors between the two times of testing. For the PSC group, there were significantly fewer TSEs at post-test than pre-test (i.e. accuracy increased) for both autocorrected errors, $t(26) = 5.282$, $p < .001$, $d = 1.014$, and non-autocorrected errors, $t(26) = 5.804$, $p < .001$, $d = 1.116$. In both cases, the effect size was large. For the MSW group, there was no significant difference between pre- and post-test scores for autocorrected errors, $t(27) = 1.695$, $p = .102$, $d = 0.321$. However, there were significantly fewer errors at post-test than pre-test for non-autocorrected TSE, with a medium-to-large effect size, $t(27) = 3.730$, $p = .001$, $d = 0.704$. Finally, for the Control group, there were no significant differences between pre- and post-test scores for either autocorrected errors, $t(26) = 1.224$, $p = .232$, $d = 0.233$, or non-autocorrected errors, $t(26) = 1.897$, $p = .069$, $d = 0.364$.

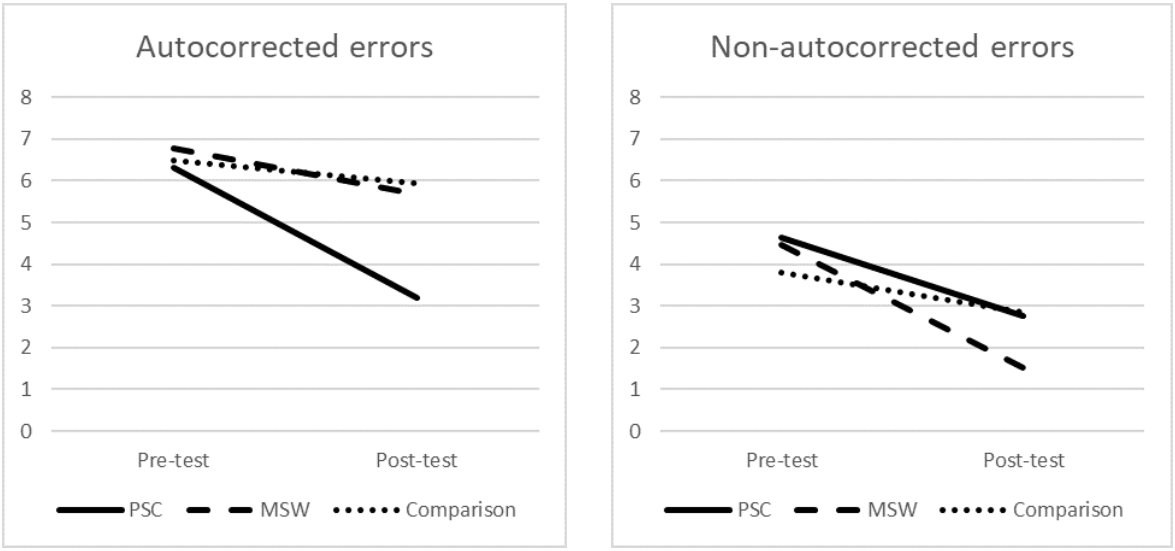
Table 4

Numbers of (a) Autocorrected and (b) Non-Autocorrected TSE Made at Pre- and Post-Test

Group (*n*)	Autocorrected errors						Non-autocorrected errors					
	PSC (27)		MSW (28)		Control (27)		PSC (27)		MSW (28)		Control (27)	
	M	***SD***	***M***	***SD***	***M***	***M***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Pre-test	6.3	3.7	6.8	3.1	6.5	3.4	4.6	2.5	4.5	2.9	3.8	2.3
Post-test	3.2	3.1	5.7	4.4	5.9	4.0	2.8	3.0	1.5	1.6	2.9	2.7

Figure 4

Mean Numbers of (a) Autocorrected TSE and (b) Non-Autocorrected TSE Made at Pre- and Post-Test



Time Demands

Our final research question asked whether there was a difference in the time taken to produce L2 text by learners using (a) the PSC, (b) the MSW spellchecker and (c) no spellchecker. This was to check whether any advantages found for any of the interventions might simply be due to more time spent on task (and thus engaging with the L2), rather than to the nature of the feedback itself.

The average time spent by participants in each group on the five transcription tasks (completed as part of the intervention) is shown in Table 5. A one-way ANOVA found no significant difference in time taken between the three groups, $F(2, 79) = 2.35$, $p = .102$. In other words, there was no evidence that writing in L2 took participants longer using any one of the software packages (PSC, MSW, Notepad) than the others.

Table 5

Average Time Spent by Participants in Each Group on the Five Transcription Tasks

	N	Mean time (m:s)	SD (m:s)
PSC	27	10:43	2:14
MSW	28	11:34	3:18
Control	27	10:05	1:55

Discussion

Accuracy of Targeted Spellings

All three groups made fewer errors at post-test than pre-test on the targeted spelling forms. For the PSC and MSW groups, the decrease in errors was statistically significant, whilst for the Control group it approached significance ($p = .051$). Assuming that our pre- and post-tests were indeed of equivalent difficulty, there

was thus a tendency for all participants to become more accurate at spelling the targeted forms over the course of the five-day intervention. This is not unexpected, since they were receiving four hours of intensive SFL tuition daily.

However, the accuracy increases were not equal across the groups. The improvement for the PSC group was greatest, with a very large effect size (d = 1.16), followed by the MSW group, with a medium effect size (d = 0.68). The effect size for the Control group was small (d = 0.34). Further, whilst the groups performed similarly at pre-test, at post-test the PSC group made significantly fewer targeted spelling errors than both the MSW and Control groups, with large effect sizes. Thus, we found evidence that the pedagogic feedback offered by the PSC was more advantageous for developing participants' spelling accuracy in new writing, compared to both (a) the feedback offered by Microsoft Word's inbuilt spellchecker and (b) no feedback.

Since we found no difference in the amount of time that participants spent on the interventions (whilst accepting the limitations of the self-timing procedure), time on task can be ruled out as a confounding variable. Thus, the advantage for the PSC group seemed genuinely to derive from the nature of the WCF they received.

Our study is consistent with previous evidence from CALL research (albeit mainly pertaining to grammar learning) that metalinguistic feedback may benefit L2 development, at least in the immediate term (see Heift & Hegelheimer, 2017). Unfortunately, due to high and selective attrition at delayed post-test, we were unable to gain insight into the longer-term impact of the metalinguistic feedback in our study, an area in which existing research remains "less conclusive and [which] requires further investigation" (Heift & Hegelheimer, 2017, p. 56).

Our findings also contribute to the growing evidence suggesting that direct, focused, synchronous WCF can be effective in improving foreign language learners' accuracy in new writing (e.g., Shintani & Aubrey, 2016). However, the feedback in Shintani and Aubrey's (2016) study was provided by teachers; its synchronous nature was argued, from a sociocultural perspective, to allow for optimal scaffolding by the teacher according to learners' individual needs. By contrast, the synchronous feedback in our study was provided automatically by computer software. This has important implications for pedagogy, because the intervention was provided at no cost to teachers' time.

Why was the PSC more effective than Microsoft Word spellchecker in our study? One reason may lie in the explicit instruction provided by the PSC in the targeted spelling patterns. This may have helped participants to develop explicit, declarative knowledge of these patterns, the first step in developing fluent spelling within a skill acquisition framework (DeKeyser, 2015). Participants could then deploy this explicit knowledge in the post-test transcription task, particularly as this was completed without time pressure. Future studies might investigate this further from a qualitative perspective, gathering data not only on participants' spelling outcomes, but also on their writing processes. For example, stimulated recall might shed light on the sources of knowledge which participants draw upon when spelling words, and the nature and extent of any self-correction.

We must acknowledge that, descriptively speaking, fewer participants had an L1 other than English in the PSC group than in the other two groups. Whilst this difference was not statistically significant, L1 background is a potentially confounding variable, and so it would be interesting to repeat our study with a more equal balance of L1s across the groups; or indeed to design a study which explicitly compared the effects of the PSC on participants with different L1s (e.g., English versus Chinese). This is because the spelling errors targeted by the PSC included some which were likely to result from L1 transfer, specifically from English (or other European languages). An example is the misspelling of the phoneme /s/ with a double <ss> (e.g., **asociación* for *asociación*), as often occurs in English and French (e.g., *association*). Nonetheless, it is highly likely (though we cannot confirm) that our Chinese participants also knew some English as an L2, which may then have influenced their spelling in Spanish (their L3).

Noticing Errors

It is interesting that participants in the MSW group showed no advantage over the Control group in the development of their spelling accuracy on TSE over the course of the intervention, despite receiving systematic, synchronous WCF on their spelling errors. In other words, we found no evidence that feedback provided by the Microsoft Word spellchecker helped participants increase their spelling accuracy any more than a 'no feedback' condition. At face value, this appears to contrast with Lin et al. (2017), who found positive effects of a generic computerized spellchecker.

The key to understanding these apparently contrasting findings may lie in the difference we observed in the current study between (a) those TSE which were autocorrected by Microsoft Word spellchecker and (b) those which were not. Recall that the PSC group showed significant decreases in the numbers of TSE in both these categories (i.e. they became more accurate), with a large effect size, whilst the Control group showed no significant difference between pre- and post-test scores in either category (i.e. their accuracy remained the same). However, the MSW group performed differently to both the other groups: Participants showed a significant decrease in the number of non-autocorrected TSE, with a large effect size, but no significant decrease in the number of autocorrected TSE. For non-autocorrected errors, Microsoft Word spellchecker underlines the erroneous form and, upon right click, provides a drop-down list of alternative forms from which the user must select the desired one. This mirrors the way errors were treated by the custom-built spellchecker in Lin et al. (2017). Thus, when we compare like with like in terms of the feedback mechanisms in that previous study and our own, the findings do not contrast after all: Similar kinds of feedback in the two studies both led to gains in spelling accuracy. By contrast, for autocorrected items, the nature of the feedback in the two studies differed substantially, and correspondingly, different outcomes were observed.

We hypothesise that the crucial difference between autocorrected and non-autocorrected errors is that the latter promote noticing errors (Schmidt, 1990**)** and noticing the gap between these errors and the correct forms (Swain, 1998). By contrast, autocorrection happens instantaneously and quite possibly without the learner's awareness—hence impeding noticing and, in turn, learning. If this hypothesis is correct, then an immediate pedagogical implication of our study is that learners should deactivate the autocorrect function in their word processing software, in order to notice their errors more and potentially learn from them.

We would nonetheless note that, even for non-autocorrected TSE, the size of the decrease in errors was smaller in the MSW group (d = 0.704) than the PSC group (d = 1.014). This may be because learners in the MSW group did not always select the correct form from the drop-down list. Alternatively, as suggested above, the PSC group may have drawn additional benefit from the explicit metalinguistic information they received. Again, further investigation of a qualitative nature could investigate these hypotheses further.

Accuracy of Non-Targeted Spellings

We also measured the accuracy of other spelling forms besides those targeted by the PSC. Between-groups comparisons found no significant advantage for any of the groups over the others in terms of increased accuracy on these other spelling errors (OSEs). However, we did find different patterns of change over time in the three groups. Participants in the MSW group showed a significant decrease in the number of OSEs (with a medium effect size), just as they did for the TSEs. This is unsurprising: Microsoft Word spellchecker does not differentiate between these two kinds of errors. The Control group, by contrast, showed no change in the numbers of OSE (again, as expected). Interestingly, however, in the PSC group, the decrease in OSEs touched the threshold of significance (p = .050), with a small-to-medium effect size (d = 0.41). We are unsure why the PSC group should have shown this tendency towards greater accuracy in spelling non-targeted forms than the Control group, given that they received no feedback on these errors. Perhaps their sensitivity to Spanish spelling patterns in general had been heightened by the feedback provided by the PSC. This merits further investigation.

Conclusions

As noted in the recent reviews by Heift (2017) and Heift and Hegelheimer (2017), the effects of automatic, computer-generated corrective feedback on the development of L2 writing is under-researched, despite the widespread use of spellcheckers as part of word processing packages. The current study contributes to addressing this gap in the literature.

Our findings show that a Pedagogic Spellchecker, designed specifically for learners of Spanish as a Foreign Language and providing error-specific, metalinguistic feedback on common spelling errors, can be an effective tool for improving L2 learners' spelling accuracy. It was found to be more effective in this respect than both (a) absence of feedback (control condition) and (b) Microsoft Word's inbuilt spellchecker. It is possible that this advantage for the Pedagogic Spellchecker arose because it developed participants' explicit knowledge of Spanish spelling patterns, which they were then able to draw on when spelling words in the untimed transcription tasks. Our findings are thus consistent with previous CALL research suggesting that metalinguistic feedback can be beneficial for L2 learning, at least in the short term.

However, in line with the limited previous research in this area, we found evidence that the generic Microsoft Word spellchecker (without metalinguistic feedback) could also promote the development of spelling accuracy to some extent—though only for those errors which it did not autocorrect. We hypothesise that this is because autocorrection—occurring instantaneously and without the learner's awareness—impedes the noticing of errors (and of the gap between these and the correct forms), in turn hindering learning. Therefore, where generic spellcheckers are used, our study suggests that learners may benefit from deactivating the autocorrect function to increase the noticing of errors and thus the ability to learn from them.

Our study also contributes to the small but growing body of evidence suggesting that synchronous written corrective feedback on spelling errors can be effective. In our study, this feedback was provided automatically by computer, with no cost to teaching and learning time.

We suggest that a larger-scale trial of the PSC is now warranted. In the current study, due to problems with extensive and selective attrition at delayed post-test, we were able to evaluate only the short-term effects of the PSC. Future research should therefore assess the durability of these effects. More detailed, process-oriented work is also required (e.g., using think-aloud or screen capture: see Hamel & Séror, 2016), to understand how learners interact with computerized spellcheckers and how this influences the development of their L2 spelling.

Acknowledgements

This research project would have not been possible without the cooperation of the Spanish summer courses at the Universidad de Castilla-La Mancha and the kind support of Juana Mª Blanco and Aranzazu Bernardo. We would also like to extend our thanks to the teachers of these courses for their positive engagement, and to all the students who participated in this study.

Notes

1. One outlier was removed from the PSC group. This participant had a pre-test score of 43 (over 1.5 times the inter-quartile range above the 75th percentile) but an unusually low post-test score (zero), suggesting some kind of data error.

References

Barry, C. (1992). Interactions between lexical and assembled spelling. In C. Sterling and C. Robinson (Eds.) *Psychology, spelling and education* (pp. 71–86). Multilingual Matters.

Berkel, A. (2004). Learning to spell in English as a second language. *International Review of Applied Linguistics in Language Teaching (IRAL), 42*(3), 239–258.

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.

Blázquez, M., & Fan, C. (2019). The efficacy of spell check packages specifically designed for second language learners of Spanish. *Pertanika Journal of Social Sciences & Humanities (JSSH), 27*(2), 847–861.

Burston, J. (1998). Antidote 98. *CALICO Journal, 16*(2), 197–212.

Burston, J. (2001). Exploiting the potential of a computer-based grammar checker in conjunction with self-monitoring strategies with advanced level students of French. *Calico Journal, 18*(3), 499–515.

Cestero Mancera, A. M. C., & Penadés Martínez, I. P. (2009). *Corpus de textos escritos para el análisis de errores de aprendices de E/LE (CORANE). CD-ROM* [Corpus of written texts for the analysis of errors committed by learners of Spanish as a Foreign Language]. Universidad de Alcalá.

Conroy, M. A. (2018). Contextual factors in second language learning in a short-term study abroad programme in Australia. *The Language Learning Journal, 46*(3), 311–328.

Corder, S. P. (1967). The significance of learner's errors. *IRAL-International Review of Applied Linguistics in Language Teaching, 5*(1-4), 161–170.

Council of Europe (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684.

DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 94–112). Routledge.

Graham, S., Morphy, P., Harris, K. R., Fink-Chorzempa, B., Saddler, B., Moran, S., & Mason, L. (2008). Teaching spelling in the primary grades: A national survey of instructional practices and adaptations. *American Educational Research Journal, 45*(3), 796–825.

Hamel, M-J., & Séror, J. (2016). Video screen capture to document and scaffold the L2 writing process. In C. Caws and M-J. Hamel (Eds.), *Language-learner computer interactions: Theory, methodology, and applications* (pp.137–162). John Benjamins.

Harrison, G. L., Goegan, L. D., Jalbert, R., McManus, K., Sinclair, K., & Spurling, J. (2016). Predictors of spelling and writing skills in first-and second-language learners. *Reading and Writing, 29*(1), 69–89.

Heift, T. (2004). Corrective feedback and learner uptake in CALL. *ReCALL, 16*(2), 416–431.

Heift, T. (2017). History and key developments in intelligent computer-assisted language learning (ICALL). In S. Thorne & S. May (Eds.), *Language, education and technology: Encyclopaedia of Language and Education* (pp. 289–300). Springer.

Heift, T., & Hegelheimer, V. (2017). Computer-assisted corrective feedback and language learning. In H. Nassaji & E. Kartchava (Eds.), *Corrective feedback in second language teaching and learning* (pp. 51–65). Routledge.

Justicia, F., Defior, S., Pelegrina, S., & Martos, F. J. (1999). The sources of error in Spanish writing. *Journal of Research in Reading*, *22*(2), 198–202.

Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The Modern Language Journal*, *99*(1), 1–18.

Lado, B., Bowden, H., Stafford, C., & Sanz. C. (2014). A fine-grained analysis of the effects of negative evidence with and without metalinguistic information in language development. *Language Teaching Research*, *18*, 320–344.

Lawley, J. (2015). New software to help EFL students self-correct their writing. *Language Learning & Technology*, *19*(1), 23–33. http://dx.doi.org/10125/44396

Lawley, J. (2016). Spelling: Computerised feedback for self-correction. *Computer Assisted Language Learning*, *29*(5), 868–880.

Lin, P. H., Liu, T. C., & Paas, F. (2017). Effects of spell checkers on English as a second language students' incidental spelling learning: A cognitive load perspective. *Reading and Writing*, *30*(7), 1501–1525.

Liu, Q., & Brown, D. (2015). Methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *Journal of Second Language Writing*, *30*, 66–81.

Nagata, N. (1993). Intelligent computer feedback for second language instruction. *Modern Language Journal*, *77*(3), 330–338.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge University Press.

Peñalver Castillo, M. (2009). *Estudios sobre ortografía y gramática del español* [Studies on Spanish grammar and spelling]. Axac.

RAE. (2014). *Diccionario de la lengua española* [Dictionary of the Spanish language] (23rd Ed.). Espasa.

Rimrott, A., & Heift, T. (2005). Language learners and generic spell checkers in CALL. *CALICO Journal*, *23*(1), 17–48.

Rodríguez, A. J., Villoria D. B., & Paredes A. F. (1997). *Las redacciones en lengua extranjera en bachillerato y COU: Investigación pedagógica* [Foreign language compositions in high school: A pedagogic research]. Universidad de Valladolid.

Rosa, E., & Leow, R. (2004). Computerized task-based exposure, explicitness and type of feedback on Spanish L2 development. *Modern Language Journal*, *88*, 192–217.

Sánchez Jiménez, D. (2010). El análisis de errores ortográficos de estudiantes filipinos en el aprendizaje de español como LE y su aplicación didáctica [Analysis of spelling errors committed by Filipino learners of Spanish and its didactic application]. In D. Isa de los Santos (Ed.), *Actas del primer congreso de español como lengua extranjera en Asia-Pacífico* [Proceedings of the first conference on Spanish as a foreign language in the Asia-Pacific] (pp. 106–126). Instituto Cervantes.

Sánchez Jiménez, D. (2013). Trascendencia de la ortografía en la clase de español como lengua extranjera y su aplicación didáctica [The importance of spelling in the Spanish as a Foreign Language classroom and its didactic application]. *Tonos Digital: Revista Electrónica de Estudios Filológicos*, *24*, 1–22.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied linguistics*, *11*(2), 129–158.

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174.

Shintani, N., & Aubrey, S. (2016). The effectiveness of synchronous and asynchronous written corrective feedback on grammatical accuracy in a computer-mediated environment. *Modern Language Journal*, *100*(1), 296–319.

Stirling, J. (2011). *Teaching spelling to English language learners.* Lulu.

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64–81). Cambridge University Press.

APPENDIX A. Attrition at Delayed Post-Test

Numbers and Pre-Test Scores of Attritors and Non-Attritors at Delayed Post-Test

Group	*N* (pre)	*N* (D-post)	*N* Attrition	% Attrition	Pre-test scores of attritors *M*	Pre-test scores of attritors *SD*	Pre-test scores of non-attritors *M*	Pre-test scores of non-attritors *SD*
PSC	35	20	15	42.9	11.9	6.2	10.0	6.8
MW	36	18	18	50.0	11.8	5.5	10.4	5.6
Control	36	18	18	50.0	13.0	5.1	8.1	4.5
Total	107	56	51	47.7	12.3	5.3	9.5	5.7

Note. Of the 107 participants who completed the pre-test, only 56 (52.3%) completed the delayed post-test. This attrition was evenly distributed across the groups: The sample sizes in each group remained roughly equal at delayed post-test (PSC: $n = 20$; MSW: $n = 18$; Control: $n = 18$). However, an independent samples t-test revealed that those participants who completed all three tests made significantly fewer TSE at time 1 ($M = 9.5$, $SD = 5.7$) than those who did not ($M = 12.3$, $SD = 5.3$), $t(105) = 2.558$, $p = .012$. Closer examination revealed that the significant difference resided specifically in the Control group. At pre-test, those Control group participants who went on to complete all three tests made on average 8.1 TSE ($SD = 4.5$), compared to 13.0 TSE ($SD = 5.1$) for those who dropped out before the delayed post-test. In other words, at baseline, the attritors were significantly less accurate on targeted spelling errors than the non-attritors, with a large effect size, $t(34) = 3.063$, $p = .004$, $d = 1.0$. In the PSC and MSW groups, no such significant difference was found. The large extent and biased nature of the attrition between pre-test and delayed post-test led us to focus our analysis on the pre-/post-test data only.

APPENDIX B. Participants' Characteristics at Pre- / Post-Test

	Gender	N	%	L1 background	N	%	L2 proficiency	N	%
Pre-test:	Male	35	32.7	English	82	76.6	AB	2	1.9
N = 107	Female	72	67.3	Chinese	23	21.5	LI	32	29.9
				French	2	1.9	I	46	43.0
							UI	27	25.2
Post-test:	Male	25	30.5	English	69	84.1	AB	0	0.0
N = 82	Female	57	69.5	Chinese	11	13.4	LI	22	26.8
				French	2	2.4	I	36	43.9
							UI	24	29.3

Note. AB = Absolute beginners (≈CEFR level A1); LI = Lower intermediate (CEFR A2); I = Intermediate (CEFR B1); UI = Upper intermediate (CEFR B2) (Council of Europe, 2020).

APPENDIX C. Sample Transcription Task (Delayed Post-Test)

Transcript of the audiorecording	Target Spelling Error
1) *Yo tengo **cuatro** bicicletas en casa.*	19

(I have four bicycles at home.)	
2) *Nosotros **mirábamos** el cielo.*	4
(We used to look at the sky.)	
3) *Se **sentirá útil** en su nuevo trabajo.*	6, 5
(He'll feel useful in his new job.)	
4) *En marzo **volverás** con un **talismán**.*	8, 7
(In March, you'll be back with a talisman.)	
5) ***Encontré** una moneda a cinco **milímetros** de la piscina.*	9, 3
(I found a coin five millimetres away from the pool.)	
6) ***Volví** **realmente** cansado de **Berlín**.*	12, 23, 13
(I came back really exhausted from Berlin.)	
7) *¿Cuándo tendremos esa **conversación**?*	10
(When will we have that conversation?)	
8) *Puede sufrir un cambio **dramático**.*	1
(He can suffer a dramatic change.)	
9) *El **escocés** no tiene **traducción**.*	14, 11
(Scottish is untranslatable.)	
10) *Nadamos en un **río** en Grecia.*	15
(We swam in a river in Greece.)	
11) *Nuestro **curso** empieza el **próximo** lunes.*	18, 26
(Our course starts next Monday.)	
12) *Muchas **veces** el socialismo crece en Europa.*	21
(Socialism often arises in Europe)	
13) *El **comercio** es bueno para la ciudad.*	16
(Commerce is good for the city.)	
14) *Su oponente no le puede **asistir**.*	17
(His opponent can't help him.)	
15) *Un **estudio** dice que **frecuentemente** la mejor terapia es la risa* (A study suggests that laughing is often the best therapy.)	20, 22
16) *Ella es **bióloga** y se llama Elena.*	2
(She is a biologist and her name is Elena.)	
17) *Su **corazón** era **buenísimo**.*	25, 24
(They had a really good heart.)	

Note. TSE are highlighted in bold, and "non-autocorrected TSE" underlined.

APPENDIX D. Target Spelling Errors and the Words Containing Them

The words containing the TSE appear in the order of presentation to participants: pre-test, treatment session 1, treatment session 2, treatment session 3, treatment session 4, treatment session 5, post-test, delayed post-test. The wording used here to describe the TSE is not the same as that used for the feedback in the PSC. Those marked with an asterisk were autocorrected by Microsoft Word.

TSE1* Spanish words ending in *-ático/a* are stressed on the third to last syllable and must carry a diacritic to indicate this.

Gramática (grammar), *informática* (computer science), *asiático* (Asian), *democrático* (democratic), *diplomático* (diplomatic), *automático* (automatic), *matemáticas* (mathemathics), *dramático* (dramatic)

TSE2* Spanish words ending in *-ólogo/a* are stressed on the third to last syllable and must carry a diacritic to indicate this.

Psicólogo (psicologist), *cardióloga* (cardiologist), *dermatóloga* (dermatologist), *geólogo* (geologist), *socióloga* (sociologist), *antropóloga* (anthropologist), *arqueólogo* (archaeologist), *biólogo* (biologist)

TSE3* Spanish words ending in *-metro* are stressed on the third to last syllable and must carry a diacritic to indicate this.

Kilómetro (kilometre), *diámetro* (diameter), *termómetro* (thermometer), *barómetro* (barometer), *perímetro* (perimeter), *parámetro* (parameter), *centímetro* (centimetre), *milímetro* (millimetre)

TSE4* In Spanish, the first person plural of the past imperfect of all verbs ends in *-ábamos* (with diacritic).

Estábamos (we were), *esperábamos* (we waited), *encontrábamos* (we found), *necesitábamos* (we needed), *pensábamos* (we though), *pasábamos* (we passed), *andábamos* (we walked), *mirábamos* (we looked)

TSE5* Spanish words ending in *-il*, which are stressed on the penultimate syllable (such as *difícil* [difficult] or *fácil* [easy]) must have a diacritic.

Fácil (easy), *imbécil* (imbecilic), *móvil* (mobile), *portátil* (portable), *frágil* (fragile), *versátil* (versatile), *difícil* (difficult), *útil* (useful)

TSE6* In Spanish, the third person singular of the future tense of all verbs ends in *-á* (with diacritic).

Tendrá (he will have), *podrá* (he will be able to), *encontrará* (he will find), *volverá* (he will return), *comprenderá* (he will understand), *vendrá* (he will come back), *pensará* (he will think), *sentirá* (he will think)

TSE7 Spanish words ending in *-an*, when these are stressed on the last syllable (e.g., *alemán* [German] or *capitán* [captain]) must have a diacritic.

Alemán (German), *imán* (magnet), *Pakistán, Afganistán, Milán, musulmán* (muslim), *capitán* (captain), *talismán* (talisman)

TSE8* Spanish words ending in *-as*, which are stressed on the last syllable such as *quizás* (maybe) or *jamás* (never), must have a diacritic.

Además (besides), *jamás* (never), *quizás* (maybe), *detrás* (behind), *atrás* (back), *compás* (compass), *sofás* (sofas), *volverás* (you will return)

TSE9 In Spanish, the first person singular of the simple future tense is stressed on the last syllable and ends in a vowel. Thus, it must have a diacritic.

Tendré (I will have), *veré* (I will see), *seré* (I will be), *vendré* (I will come back), *estaré* (I will be), *sabré* (I will know), *iré* (I will go), *encontraré* (I will find)

TSE10* In Spanish, almost all words ending in *-ion* must have a diacritic on the *-o-*, because they are stressed on the last syllable.

Habitación (room), *atención* (attention), *sensación* (sensation), *ocasión* (occasion), *información* (information), *expresión* (expression), *acción* (action), *conversación* (conversation)

TSE11 Spanish words ending with the suffix *-cion* are written with *-cc-* when there is at least a word in the given word family that includes the consonant group *-ct-* such as *director* (director) → *dirección* (management).

Sección (section), *reacción* (reaction), *satisfacción* (satisfaction), *dirección* (direction, address, management), *elección* (election), *destrucción* (destruction), *acción* (action), *traducción* (translation)

TSE12 In Spanish, the first person singular of the simple past tense is stressed on the last syllable and ends in a vowel. Thus, it must have a diacritic.

Escribí (I wrote), *conocí* (I knew), *salí* (I went out), *creí* (I believed), *respondí* (I answered), *sentí* (I felt), *comí* (I ate), *volví* (I returned)

TSE13 Spanish words ending in *-in*, which are stressed on the last syllable such as *jardín* (garden) or *Berlín*, must have a diacritic.

Jardín (garden), *delfín* (dolphin), *Medellín, Dublín, violín, latín, calcetín* (sock), *Berlín*

TSE14* In Spanish, nationalities and languages ending in *-es* are stressed on the last syllable and must therefore have a diacritic.

Francés (French), *finlandés* (Finnish), *japonés* (Japanese), *portugués* (Portuguese), *irlandés* (Irish), *holandés* (Dutch), *inglés* (English), *escocés* (Scottish)

TSE15 In Spanish, words ending in *-io* and *-ia* must have a diacritic on the *-i-* when these vowels are pronounced in different syllables.

Frío (cold), *mía* (mine), *mío* (mine), *tía* (aunt), *día* (day), *tío* (uncle), *fría* (cold), *río* (river)

TSE16 In Spanish, there are no words written with *-mm-*.

Inmediatamente (immediately), *comunicación* (communication), *inmortal* (immortal), *programa* (programme), *inmoral* (immoral), *comando* (commando), *común* (common), *comercio* (commerce)

TSE17 In Spanish, there are no words written with *-ss-*.

Acceso (access), *agresión* (aggression), *depresión* (depression), *misil* (missil), *asesinato* (assassination), *asalto* (assault), *posible* (possible), *asistir* (assist)

TSE18 The only word in Spanish that contains the sequence *-ou-* is *estadounidense* (American).

Grupo (group), *duda* (doubt), *acústico* (acoustic), *anuncio* (announce), *ducha* (shower), *curioso* (curious), *turista* (tourist), *curso* (course)

TSE19 In Spanish, the syllable /kwa/ is written *-cua-* with *c* not with *q*.

Cualificado (qualified), *ecuador* (equator), *cuantitativo* (quantitative), *acuario* (aquarium), *cual* (who, what), *cuadrado* (squared), *cuando* (when), *cuatro* (four)

TSE20 No word in Spanish begins with an *s-* followed by a consonant.

Específico (specific), *estación* (station), *espíritu* (spirit), *escena* (scene), *espacio* (space), *estado* (state), *especial* (special), *estudio* (study)

TSE21 In Spanish, when a word in singular form ends in *-z*, it is transformed into the plural form by changing the *z* with a *c* and then adding *-es*.

Felices (happy), *actrices* (actresses), *peces* (fishes), *dieces* (tens), *voces* (voices) , *cruces* (crosses), *lápices* (pencils), *veces* (times)

TSE22* In Spanish, the suffix *-mente* should be added to the feminine form of an adjective. However, neutral adjectives ending in *-e* are transformed into an adverb by just adding the suffix *-mente*.

Probablemente (probably), *evidentemente* (evidently), *suficientemente* (sufficiently), *posiblemente* (possibly), *constantemente* (constantly), *terriblemente* (terribly), *simplemente* (simply), *frecuentemente* (frequently)

TSE23 In Spanish, neutral adjectives ending in *-l* are transformed into an adverb adding just *-mente*.

Finalmente (finally), *totalmente* (totally), *facilmente* (easily), *personalmente* (personally), *principalmente* (principally), *normalmente* (normally), *generalmente* (generally), *realmente* (really)

TSE24* Spanish words ending in the strings of letters *-ísimo/a* are stressed on the third to the last syllable and, therefore, must have a diacritic.

Altísimo (very tall), *grandísima* (very big), *larguísimos* (very long), *malísimos* (very bad), *muchísimos* (a lot), *rarísimo* (very strange), *muchísima* (a lot), *buenísimo* (very good)

TSE25* Spanish words ending in *-on*, which are stressed on the last syllable such as *camión* (truck) or *melón* (melon), must have a diacritic.

Balón (ball), *campeón* (champion), *botón* (button), *marrón* (brown), *sillón* (armchair), *melón* (melon), *razón* (reason), *corazón* (heart)

TSE26* Spanish words which are stressed on the antepenultimate syllable must have a diacritic.

Miércoles (Wednesday), *bolígrafo* (pen), *cámara* (camera), *química* (chemistry), *teléfono* (telephone), *rápido* (fast), *sábado* (Saturday), *próximo* (next)

About the Authors

Miguel Blázquez is Assistant Professorial Fellow and Spanish Section Coordinator at the University of the Philippines Diliman Department of European Languages. He holds a PhD in Spanish Philology and an MSc in Applied Linguistics and Second Language Acquisition. His research interests include L2 writing, WCF, corpus linguistics, and CALL.

E-mail: mblazquez@up.edu.ph

Robert Woore is Associate Professor in Applied Linguistics at Oxford University Department of Education. He works on the department's Master's, Doctoral and Initial Teacher Education programmes. Formerly, he taught French and German at secondary school. His main research focus is the learning and teaching of reading in a foreign language.

E-mail: robert.woore@education.ox.ac.uk