Feedback Treatments, Writing Tasks, and Accuracy Measures: A Critical Review of Research on Written Corrective Feedback

Nang Kham Thi
University of Szeged, Hungary
<nang.kham.thi@edu.u-szeged.hu>

Marianne Nikolov
University of Pécs, Hungary
University of Szeged, Hungary
<nikolov.marianne@pte.hu>

Abstract

With a considerable emphasis on the role of feedback in L2 writing, the effectiveness of written corrective feedback (WCF) has been investigated extensively over the past 25 years. Conflicting findings have been reported regarding the efficacy of WCF in developing learners' written accuracy. This paper provides an in-depth analysis of research on three key variables impacting the effects of WCF: feedback treatments; writing tasks with more/less cognitive and reasoning demands; and L2 written accuracy measures. Forty-two primary studies published between 2000 and 2020 were retrieved and coded following PRISMA guidelines. Results revealed variations in both feedback treatments and written accuracy measures with distinctive advantages and pitfalls. Further divergent issues concern the application of different genres of writing tasks that demand learners' various cognitive and linguistic efforts. These variations make it difficult to compare results across empirical studies. The findings contribute to a better understanding of why and how uniform criteria for selection of writing tasks and accuracy measures can ensure the comparability of studies.

Keywords: *written corrective feedback, feedback treatments, grammatical accuracy measures, second language writing*

Correcting learner errors through written feedback has long been part of pedagogical traditions that writing teachers practice and is of widespread interest of second language (L2) writing research. Due to the multifaceted nature of the construct of writing, feedback on students' writing includes "a wide variety of responses and may contain information regarding the accuracy, communicative success, or content of learner utterances or discourse" (Leeman, 2010, p. 112). Pedagogically, feedback links assessment to teaching and learning. It reflects information about learners' actual performance and guidance on

future learning goals. Based on that premise, considerable attention has been paid to the role of feedback in writing classrooms. Written feedback plays a part in helping learners improve writing accuracy, task achievement, and organization.

Many studies have examined the role of written corrective feedback (WCF) in writing classrooms (Ferris, 2006; Truscott & Hsu, 2008), comparing various feedback types (Bitchener, 2008; Ellis et. al., 2008). The findings have been synthesized in a number of meta-analyses and review papers. The first meta-analysis by Russell and Spada (2006) provided support for the effectiveness of corrective feedback for L2 grammar learning. Subsequently, Hyland and Hyland (2006) conducted an extensive narrative review, stressing the diversity in student populations, types of writing, feedback practices, and research designs. Truscott (2007) looked at the effect of error correction in twelve published studies and concluded that correction had a harmful effect on students' ability to write accurately. In 2011, another meta-analysis by Biber et al. examined 25 published studies from 1982 to 2007 and suggested that feedback result in accuracy gains in writing development. Kang and Han (2015) later identified 21 primary studies (from 1980 to 2013) and examined whether WCF helped to improve grammatical accuracy. They concluded that WCF led to greater grammatical accuracy in L2 writing. Unlike the earlier meta analyses, Liu and Brown (2015) reviewed 32 published studies and twelve dissertations and identified methodological limitations. More recently, Sia and Cheung (2017) conducted a qualitative synthesis of 68 empirical studies published in journals from 2006 to 2016, shedding light on the role of individual differences in the effectiveness of WCF. Karim and Nassaji (2019) presented a critical synthesis of research on WCF and its effects on L2 learning over the past four decades. Li and Vuono (2019) specifically reviewed 25 years of research on WCF in *System*. The most recent meta-analysis by Lim and Renandya (2020) included 33 studies and five unpublished Ph.D. and Master's dissertations published between 2001 and 2019. The results of these reviews and meta-analyses provided evidence that WCF has the potential to improve L2 written grammatical accuracy and is conducive to writing development, except the findings from the meta-analytic study by Truscott (2007).

Nonetheless, issues pertaining to the extent to which WCF is helpful and which feedback treatments (e.g., implicit versus explicit) bring long-term effectiveness have been widely debated. Convincing answers have remained difficult to obtain (Ferris, 2004; Truscott, 2007). Truscott (1996) sparked considerable discussion in the literature, arguing that feedback provision has little or no contribution to the development of accuracy in writing (Truscott, 1996, 2004, 2007, 2010; Truscott & Hsu, 2008). Despite Truscott's (2007) call for the abandonment of error correction, subsequent and recent studies have supported the overall benefits of WCF (e.g., Ferris, 1999, 2004; Karim & Nassaji, 2018; Van Beuningen et al., 2012). However, differences in feedback treatments, targeted linguistic features, and heterogeneity of participants probably impact the efficacy of WCF.

Furthermore, methodological limitations with regard to the use of a wide range of accuracy measures render it difficult to ensure comparability across studies (for review see Liu & Brown, 2015). Other limitations concern the effects of different genres of writing tasks, a significant moderator variable, contributing to variance in effect size (for a review see Kang & Han, 2015). Thus, the present study attempts to review, identify, and synthesize research on WCF. This review aims to add information to previous syntheses about the effectiveness of WCF research illuminating the importance of consistency in feedback treatments and genres of writing tasks. Hopefully, this will make

comparability of the findings possible. Drawing insights from this review, future researchers can enhance their understanding of advantages and pitfalls of accuracy measures to make informed choices.

Theoretical Background

Theoretical Understandings of the Role of Written Corrective Feedback

Although studies in WCF research are empirically motivated due to its direct relevance to the work of teachers in classrooms, they tend to be embedded in theories underpinning the potential contribution of WCF to L2 development (Bitchener, 2012; Polio, 2012; Zhang, 2021). Skill-acquisition theories (DeKeyser, 2007) proposed that accuracy is a function of practice and that explicit instruction and extensive practice are preconditions to converting declarative to procedural knowledge. In line with these claims, WCF aims to enable learners to store and retrieve declarative knowledge, which can include explicit knowledge of the target language. For example, Evans et al. (2011) and Hartshorn and Evans (2015) acknowledged the importance of practice and feedback to facilitate greater automatization within the framework of skill-acquisition theory. Another theoretical stance underlying WCF is the noticing hypothesis (Schmidt, 2010). Giving conscious attention to linguistic forms and noticing negative evidence, WCF is meant to help learners notice the gaps between their interlanguage (i.e., a natural language produced by L2 learners) and the target language. Yet another theoretical argument supporting WCF concerns the interaction hypothesis. Long's (1980) interactionist approach was originally designed with oral communication in mind. It is also pertinent to WCF as the two types of input (i.e., positive and negative evidence) are of equal relevance for both oral and written feedback. Some recent studies (e.g., Frear & Chiu, 2015) were framed along the interaction hypothesis, suggesting that WCF provides opportunities for the provision of negative evidence.

Feedback Treatments in Written Corrective Feedback

In L2 writing, scholars and teachers have stressed the importance of written feedback in developing students' writing abilities. Based on the dichotomy between feedback on form and content, written feedback could be classified into corrective and non-corrective feedback (Luo & Liu, 2017). Corrective feedback promotes the learning of the target language by providing negative evidence and non-corrective feedback scaffolds writing in aspects of content, organization, linguistic performance, and format. In other words, corrective feedback focuses on developing students' accuracy, whereas non-corrective feedback provides commentary to rhetorical and content issues (Goldstein, 2006, 2004). Compared to research investigating the influence of written commentary feedback on students' revision and future texts, there has been increased interest to determine the role of different types of WCF in L2 writing. As suggested by Storch and Wigglesworth (2010), corrective feedback can be differentiated based on "its directness which ranges from direct (e.g., writing the correct form above the incorrect form) to indirect (e.g., using editing symbols to signal an error)" (p.304). With reference to empirical studies on WCF, Ellis (2009b) identified three major strategies for providing feedback: direct, indirect, and metalinguistic feedback. Direct feedback is given through the provision of the correct target language form. Indirect feedback is provided implicitly, by an indication that an error has been committed. Metalinguistic feedback provides students with some form of explicit explanations about errors (Ellis, 2009b). Accordingly, empirical studies have

investigated the facilitative role of feedback either by comparing specific feedback strategies over no-feedback conditions (e.g., Kurzer, 2018; Truscott & Hsu, 2008) or by comparing the relative effectiveness of two or more feedback strategies (e.g., Mirzaii & Aliabadi, 2013; Riazantseva, 2012).

Further distinctions in feedback treatments concern the scope of feedback (comprehensive versus focused) – i.e. "the amount of WCF teachers should give to students – whether to respond to all written errors or to respond to them in a selective or focused manner" (Mao & Lee, 2020, p.1). Some earlier studies (e.g., Han & Hyland, 2015; Lee, 2004, 2005) investigated the nature of teachers' written feedback practices in classrooms and found that both teachers and students preferred comprehensive error correction. However, studies in WCF research have put emphasis on the utilization of focused feedback (e.g., Benson & DeKeyser, 2018; Stefanou & Révész, 2015), suggesting that responding to errors in a focused manner is more beneficial than responding to all errors in an unfocused manner. Other recent studies have taken a comprehensive approach (e.g., Bonilla Lopez et. al., 2017; Van Beuningen et. al., 2012) preferring feedback on diverse errors rather than on errors of a single type.

The Role of Writing Tasks in L2 Writing

As Ellis (2009c) posited, the primary focus of a task in language learning should be on meaning with a clearly defined outcome and learners should rely on their linguistic or non-linguistic resources to complete a task. Tasks can either be unfocused or focused (p. 223) based on a distinction between whether a task requires learners to use language in general or use specific linguistic feature. In L2 writing, both unfocused (e.g., essay) or focused writing tasks (e.g., grammatical structure) are used to assess learners' L2 writing proficiency. In the case of unfocused writing tasks, understanding how task demands impact variations in the quality and quantity of L2 writing plays a part in eliciting specific levels of L2 performance, as "tasks provide a context for negotiating and comprehending the meaning of language provided in task input" ,and "tasks provide opportunities for uptake of (implicit or explicit) corrective feedback on a participant's production" (Robinson, 2011, p. 4). Adopting Robinson's cognition hypothesis, Kuiken and Vedder (2008), for instance, examined the effect of a writing prompt on complexity and accuracy. They concluded that texts written in response to more cognitively demanding tasks turned out to be more accurate, with lower error ratios per T-unit, than those of cognitively less demanding tasks. These findings informed L2 researchers how interactions between the genres and cognitive demands of the writing tasks impact students' writing accuracy.

More specifically in WCF research, how cognitive demands of writing tasks affect learners' accuracy has received relatively scant attention. For example, Riazantseva (2012) examined the effects of WCF along three outcome measures of writing performance: in-class essays, in-class summaries, and at-home summaries which differed in terms of cognitive and linguistic demands. The findings suggested that these outcome measures produced different estimates of L2 writing accuracy.

Task-related factors including task types and task complexity are supposed to impact the score reliability of students' writing in both high-stakes and classroom assessment contexts (Liu & Huang, 2020). In terms of genres of writing tasks, empirical investigations (e.g., Kuiken & Vedder, 2008; Polio & Yoon, 2018; Yoon & Polio, 2017) provided evidence that different genres have different communicative or functional requirements that may result in different language use. For example, Yoon and Polio

(2017) posited that more complex language can occur in argumentative essays because they have higher reasoning demands than narrative essays. Similarly, Polio and Yoon (2018) found that the functional requirements for narrative and argumentative writing are different; thus, the two genres require different language. These findings help deepen our understanding about the impact of diverse writing tasks on learners' linguistic performance (e.g., accuracy and complexity).

The Role of Linguistic Accuracy in L2 Writing

Linguistic accuracy which has been defined as "the ability to be free from errors while using language to communicate in either writing or speech" (Wolfe-Quintero, Inagaki, & Kim, 1998, p. 33) is a relevant construct for research in L2 writing assessment and pedagogy. As noted by Ferris (2006), "accuracy in writing matters to academic and professional audiences" (p. 81). Polio and Shea (2014) enumerated five reasons of measuring accuracy in L2 writing research: to investigate (a) the effects of WCF, (b) the effects of planning, (c) the effect of task complexity, (d) the difference between individual and collaborative writing, and (e) change over time. Specifically to WCF research, attention given to linguistic accuracy in writing classes has been a reason of measuring accuracy gains and both L2 writing teachers and students have agreed that written accuracy is expected in academic writing. Lee (2008), for instance, found that 94.1% of teacher feedback focused on form, 3.8% on content, 0.4% on organization, and 1.7% on other aspects when investigating the feedback practices of teachers in Hong Kong secondary English classes. In a similar vein, the review by Liu and Brown (2015) reported that 36% of empirical studies provided feedback solely on grammar, 18% included feedback focusing on both grammatical and lexical errors, and 27% provided extensive feedback, whereas other studies did not specify the focus of feedback. Other possible reasons of the providing feedback on language-related errors rather than on content-related issues may be related to their beliefs (what teachers assume students can deal with) and students' level of proficiency. They may also be impacted by contextual factors including time constraints, teacher workload, and large class sizes (Mao & Crosthwaite, 2019).

In the WCF literature, students' accuracy has been the key dependent measure used to assess the effects of feedback. As Nicolas-Conesa et al. (2019) stated, a distinction was made between feedback for accuracy and feedback for acquisition with reference to a dichotomy between uptake (i.e., errors successfully corrected in rewritten texts) and retention (i.e., reduction in error-making over time). In other words, whereas feedback for accuracy concerns how the provision of feedback helps improve learners' accuracy shortly after processing it, feedback for acquisition favors "long-term language learning by involving students in feedback processing, detection of errors, self-reflection on errors, and new output" (p. 849). Though earlier studies (e.g., Ferris & Roberts, 2001) measured accuracy gains by comparing the accuracy of students' first drafts and revised texts, Truscott and Hsu (2008) claimed that accuracy gains in learners' rewritten texts failed to provide evidence that feedback provision is beneficial for acquisition. Bearing this claim in mind, recent empirical studies (e.g., Sheen et. al., 2009; Van Beuningen et al., 2012) included new pieces of writing in their research designs and compared outcome accuracy developments in both revised and new texts.

As for accuracy measures, Polio and Shea (2014) investigated the current measures of linguistic accuracy used in L2 writing research and found that holistic scales, error-free

units, number of errors, number of specific error types, and measures that take into account error severity were the primary measures. With these accuracy measures in mind, WCF studies have applied different written accuracy measures in line with their research aims, rendering it impossible to compare findings across studies.

Method

This section overviews how data searches were carried out along five criteria for inclusion and discusses the data analysis procedure. We followed the guidelines of the Preferred Items for Systematic Reviews and Meta-Analysis (PRISMA) statement (Moher et al., 2009) to ensure that our review is systematic. First, inclusion/exclusion criteria were established, and relevant studies were identified through electronic and hand searches. Then, the coding scheme was developed drawing on a framework for analyzing error correction studies (Ferris, 2003). Lastly, detailed analyses were conducted followed by synthesizing and interpreting the findings.

Literature Search: Identifying Primary Studies

Peer-reviewed journal articles discussing the effects of WCF on L2 written accuracy development were retrieved from the electronic databases of the Educational Resources Information Center (ERIC), ScienceDirect, Scopus, and Google Scholar using the following key terms: *Written corrective feedback*, *comprehensive corrective feedback*, and *feedback in second or foreign language writing*. In addition to electronic searches, a hand search in the key journals of L2 writing was conducted to ensure that empirical studies were identified. Journals included *Journal of Second Language Writing, Language Teaching Research, System, Assessing Writing, TESOL Quarterly, Modern Language Journal, Language Learning,* and *Language Teaching*.

The literature search covered studies published from 2000 to 2020 and the initial search in the databases resulted in a pool of thousands of journal articles, book chapters, books, and review articles. Due to a bulk of empirical studies examining the effects of WCF on linguistic accuracy, we included primary studies along these criteria: (a) WCF must be the focus of the study, (b) it must explicitly describe methodological considerations, (c) it must consider text samples that include the production of either revised or new texts, (d) it must utilize unfocused writing tasks in which students are allowed to produce language with relatively few constraints and with meaningful communication (Ellis, 2009c; Norris & Ortega, 2000) (e) WCF must be provided by a teacher and/or a researcher, and (f) study must be written in English.

A study was excluded if it (a) focused mainly on learners' beliefs and engagement with WCF (e.g., Han, 2017; Han & Hyland, 2015), (b) considered the effectiveness of peer feedback or automated feedback (e.g., Luo & Liu, 2017), (c) focused exclusively on learners' perceptions and how individual differences mediate the effectiveness of WCF (e.g., Park, Song, & Shin, 2016) without assessing accuracy gains, and (d) concerned how learners process written feedback (e.g., Kim & Bowles, 2019). The database search delivered more than 20,000 references and hand searches added 50 studies. After removing duplicates and articles that did not satisfy the inclusion criteria, the review yielded a sample of $n = 42$ studies which were then judged for their quality and relevance.

Data Analysis Procedure

Data analysis was conducted in three iterative phases. First, the 42 studies were carefully read to identify their theoretical perspectives. Second, different aspects of feedback treatments, writing tasks, and accuracy measures were reviewed to ensure that they were not left out in the coding scheme. Third, a draft coding scheme was developed to organize all relevant information related to each study. The development of the coding scheme was guided by a framework for analyzing error correction studies (Ferris, 2003). Our coding scheme comprises three categories: (1) feedback treatments, (2) writing tasks, and (3) measures for linguistic accuracy (see Table 1). These were further divided into 13 sub-categories after identifying contributing variables in the data set. After the coding scheme was established, the selected studies were categorized in the scheme.

Table 1. Coding Scheme of Three Categories and 13 Sub-categories.

Categories	Variables
Feedback treatments	Feedback type: direct feedback, indirect feedback, comparison of direct and indirect feedback, comparison of direct and metalinguistic feedback, comparison of different types of indirect feedback, and etc. Feedback focus: article, preposition, past tense, hypothetical condition, multiple sentence level issues (organization, paragraphing, cohesion, relevance), a wide range of grammatical errors, and a wide range of grammatical and lexical errors, verb tense choice Feedback mode: teacher, researcher, teacher-researcher, not stated Feedback scope: focused feedback, unfocused feedback, comparison of focused and unfocused feedback, not stated
Writing tasks	Genre: narrative, argumentative, autobiography, picture-composition, dictogloss, and etc. Time: timed writing, untimed writing, both timed and untimed writing, not stated Provision of prompts: yes (pictures, pictures plus prompts, reading text, listening task), no, not applicable, not stated No. of writing during the intervention: open-ended Length of writing: open-ended Revision: yes, no New writing tasks: yes, no
Measures of linguistic accuracy	Measures: error ratio, error-free clause ratio, error-free T-unit ratio, holistic scoring, obligatory occasion analysis, and etc. Reliability: inter-, intra-, both inter-and-intra, no

Results and Discussion

Surface Properties of the 42 Selected Empirical Studies

Examining the distribution of the studies' characteristics revealed that a large proportion of studies (93%) targeted adult learners enrolled either in general English classes or in academic writing classes. Seven percent targeted teenagers between 11 and 17. In relation to participants' L2 proficiency level, approximately half of the studies (45%) recruited learners at high and low intermediate level of L2 proficiency. Some researchers (e.g., Stefanou & Révész, 2015) posited that recruiting participants with intermediate

proficiency increased the comparability of their research to previous studies which targeted learners at intermediate levels. Some studies (e.g., Bonilla Lopez et al., 2017) recruited students with low and high proficiency levels and found that WCF effectively enhanced students' immediate grammatical accuracy and accuracy improvement regardless of their L2 proficiency.

Sample sizes in the data set ranged from 27 to 325 with single or multiple language backgrounds. In particular, 59% of the data set recruited participants from multiple language groups, whereas 36% came from single groups; 5% did not report the students' language background. Most of the 42 studies (88%) were conducted in educational contexts where English was taught either as a second or a foreign language. A further imbalance was found with respect to the studies' language contexts. Though WCF research was conducted worldwide, 18 studies were conducted in US contexts and 24 studies were done in EFL contexts such as Japan, Korea, Laos, and Spain. Furthermore, 69 per cent were done in university settings, with considerably fewer studies (9.5%) (e.g., Fazio, 2001; Van Beuningen et al., 2008; Van Beuningen, De Jong, & Kuiken, 2012) in other educational settings including elementary, secondary, and high schools. Thus, most studies examined how university students respond to WCF and provided less evidence on how younger learners act upon teacher feedback.

Feedback-related Features in WCF Studies

As shown in Table 2, three feedback types were identified in the 42 studies: (a) direct feedback, (b) indirect feedback, and (c) metalinguistic explanation. These studies investigated how WCF functions through comparing feedback with no feedback conditions or by comparing different feedback strategies. The most frequently applied design compared the differential effectiveness of direct and metalinguistic feedback (29%), followed by indirect feedback (21%), the comparison of direct and indirect feedback (14%), and different direct feedback types (14%). As feedback types are regarded crucial in influencing the effects of WCF, controversies relating to this factor need special attention.

We found that combined feedback strategies complicate how a single feedback type functions. For example, direct feedback alone can be turned into many feedback types such as direct focused feedback, direct unfocused feedback, direct feedback with metalinguistic explanation, direct feedback with revision, and direct feedback followed by individual conference. These variations render it difficult to compare and generalize the findings across studies. Discrepancies in feedback types still exist even when a similar feedback type is provided. For example, though studies employed metalinguistic feedback in a similar vein, the students in Shintani and Ellis's (2013) study received handouts with the rule of the targeted linguistic structures, whereas participants in Benson and DeKeyser's (2018) study had their errors marked and received consistently worded metalinguistic comments in the form of brief grammar rules on their Microsoft Word documents. These trivial variances in how similar feedback is provided most probably contributed to some degrees of the (in)effectiveness of WCF.

Table 2. Feedback-related Features.

Variables	Levels	No. of studies	%
Feedback type	Comparison of direct and metalinguistic explanation	12	29
	Indirect feedback	9	21
	Comparison of direct and indirect feedback	6	14
	Comparison of different direct feedback types	6	14
	Comparison of direct, indirect, and metalinguistic explanation	4	10
	Direct feedback	2	5
	Comparison of different indirect feedback types	2	5
	Indirect feedback followed by direct feedback	1	2
Feedback focus	A wide range of grammatical features	15	32
	A wide range of grammatical and lexical errors	9	19
	Multiple sentence level issues (organization, paragraphing, cohesion, relevance)	9	19
	Article	7	15
	Past tense	3	7
	Hypothetical condition	2	4
	Preposition	1	2
	Verb tense choice	1	2
Feedback mode	Researcher	18	43
	Teacher	13	31
	Teacher-Researcher	8	19
	Not stated	3	7
Feedback scope	Comprehensive feedback	21	50
	Focused feedback	17	40
	Comparison of focused and unfocused feedback	3	7
	Not stated	1	2

With respect to feedback focus, 32% of the data set (e.g., Benson & DeKeyser, 2018; Bitchener et al., 2005; Van Beuningen et al., 2008) provided feedback on linguistic aspects of writing such as verb tense, verb form, articles, singular-plural, and subject-verb agreement. However, the linguistic features targeted in the studies were not well-balanced: seven studies focused solely on two functions of indefinite and definite English article systems. We found that 19% provided feedback on grammatical and lexical errors (e.g., Mawlawi Diab, 2015; Riazantseva, 2012) or on form and multiple sentence-level issues such as organization, paragraphing, cohesion, and relevance (e.g., Ashwell, 2000; Chandler, 2003; Evans et al., 2010; Fazio, 2001; Ferris, 2006). Despite the prevalent emphasis on how WCF helps improve linguistic accuracy of L2 texts, we found that few studies (e.g., Hartshorn & Evans, 2015; Hartshorn et al., 2010; Van Beuningen et al., 2012) examined how WCF dealt with other aspects of writing (e.g., writing fluency, writing complexity, lexical diversity, and rhetorical appropriateness) in addition to measuring accuracy.

In connection with feedback mode, researchers were the predominant source of feedback (e.g., Benson & DeKeyser, 2018; Bonilla Lopez et al., 2017, 2018; Sheen et al., 2009; Stefanou & Révész, 2015) due to logistical and methodological reasons. In particular,

43% of the studies stated that researchers provided feedback to ensure the consistency and to avoid influencing the results, whereas teachers provided feedback in 31% of the studies (e.g., Evans et al., 2010; Ferris, 2006; Rummel & Bitchener, 2015; Vyatkina, 2010). In the latter cases, there are additional variations: either classroom teachers or other teachers who did not teach the target classes provided the WCF. In Fazio's (2001) study, for instance, a francophone elementary teacher who was not one of the classroom teachers offered feedback on students' texts to minimize variability in feedback quality due to teacher effects and to strengthen the research design. In contrast, in other studies (e.g., Ferris, 2006; Ferris & Roberts, 2001; Hartshorn & Evans, 2015; Vyatkina, 2010) multiple teachers gave feedback in their intact classes. Overall, 19% of the studies claimed that the instructor was one of the researchers. Though no previous studies have considered the difference between feedback effects depending on the source of feedback (i.e., either from researcher or teacher), it is an important variable: students' motivation and their engagement with feedback may be higher if they receive it from their teachers, whereas they may attend to feedback less if they know it was provided by a researcher. Moreover, the quality and quantity of teachers' feedback provided may be more tuned to their students' needs. Thus, Liu and Brown (2015) suggested that training should be provided when teachers provide WCF in a study to better control the variations resulting from the source of feedback.

In terms of feedback scope, half of the selected studies (e.g., Ashwell, 2000; Chandler, 2003; Truscott & Hsu, 2008; Van Beuningen et al., 2008) used comprehensive feedback, whereas 40% (e.g., Mawlawi Diab, 2015; Shintani & Ellis, 2013, 2015; Shintani et al., 2014) applied focused feedback. At this point, issues related to focused and unfocused approach are worth discussing explicitly, as both have distinctive strengths and drawbacks. The focused approach has the advantage of yielding a greater effect of WCF within a time frame. However, this approach reduces ecological validity, as it does not seem to represent feedback practices used in writing classes (Van Beuningen et al., 2012). Drawbacks include the difficulty in deciding areas that learners find difficult and if a single error type is considered, it is unnecessary to use direct testing of writing. Instead, grammar exercises focusing on specific language features can be utilized for this purpose. Moreover, the absence of obligatory occasions of the targeted error types in students' subsequent writing might impact the process of measuring accuracy gains.

Unlike focused feedback, comprehensive feedback tends to be more compatible with classroom practices due to unlimited foci on error categories regardless of whether they are related to form- or meaning-focused aspects of writing. However, some issues in comprehensive feedback studies resulted from a wide range of error types that they targeted. For example, as comprehensive feedback targets many aspects of writing, challenges relating to how to deal with almost all aspects of writing and how to offer consistent feedback need to be considered. Due to its unlimited feedback scope, it is more time-consuming compared to the focused approach.

The review revealed a new trend of research, (7%, Ellis et al., 2008; Frear & Chiu, 2015; Sheen et al., 2009) comparing the impact of focused and unfocused feedback. A key question of whether focused or unfocused feedback leads to higher accuracy gains is still an open one in need of future investigations, as the findings revealed conflicting results. For example, Ellis et al.'s (2008) study demonstrated that the direct unfocused group improved in terms of accuracy compared to the direct focused group initially, but the

Table 3. Writing Task-related Features.

Variables	Levels	No. of studies	%
Genres	Narrative writing	12	29
	Paragraph writing (during treatment) plus opinion-led essays (pre-and post-test)	6	14
	Picture description	6	14
	Opinion essay	3	7
	Email writing	3	7
	Autobiographical essay	2	5
	Dictogloss	2	5
	Argumentative essay	2	5
	Letter (e.g., informal, job application)	2	5
	Journal writing	1	2
	Persuasive essay	1	2
	Essay and summary	1	2
	Not stated	1	2
Provision of prompts	Yes (pictures/prompts/reading text/ listening task)	22	52
	No	12	29
	Not stated	7	17
	Not applicable	1	2
Time constraints	Timed writing	33	79
	Not stated	5	12
	Untimed writing	3	7
	Both timed and untimed writing	1	2
Length of writing	30 minutes	11	26
	10 minutes (treatments) + 30 minutes (pre- and post-test)	6	14
	Not stated	6	14
	20 minutes	4	10
	50 minutes	3	7
	No limits	2	5
	70 words	1	2
	100 words	1	2
	500 words	1	2
	5 pages	1	2
	8 pages	1	2
	12 minutes	1	2
	15 to 20 minutes	1	2
	25 minutes	1	2
	45 minutes	1	2
	1 hour (timed) * untimed writing included	1	2
No. of writings	Three	11	26
	Four	9	21
	Five	6	14
	Many short texts	6	14
	Two	5	12
	Not stated	3	7
	One	2	5
Revision	Yes	34	81
	No	8	19
New writing task	Yes	39	93
	No	3	7

focused group continued to improve in the long run. A study conducted by Sheen et al. (2009) also found that unfocused feedback is of limited pedagogical value when

compared to direct focused feedback. However, this was not the case in Frear and Chiu's (2015) inquiry which claimed that both focused and unfocused WCF groups outperformed the control groups on immediate and delayed post-tests.

Writing-task-related Features in WCF Studies

As for the impact of different genres of writing tasks, we found that among the range of task types, narrative writing (29%) and paragraph writing with a variety of genres (14%) were the most predominant (see Table 3). Despite using unfocused writing tasks in the data set, inconsistencies emerged from diverse linguistic and cognitive efforts that different genres demand.

Even when similar genres are used, noticeable differences relate to writing prompts. Concerning this aspect, two distinctions can be made: whether the prompts are provided and how they are operationalized. Overall, 22 studies (52%) (e.g., Bitchener, 2008; Bitchener & Knoch, 2008; Truscott & Hsu, 2008; Van Beuningen et al., 2008) offered writing prompts (either word prompts or picture prompts or both) to assist L2 writers with unfamiliar vocabulary items, whereas the other 12 studies (29%) did not make any claims about the provision of writing prompts. The other distinction concerns how writing prompts were operationalized: 18 studies offered four to six picture series (e.g., Benson & DeKeyser, 2018; Karim & Nassaji, 2018; Shintani & Ellis, 2013) to trigger L2 writers' ideas. The writing tasks in Benson and DeKeyser (2018) and Sheen et al. (2009) comprised word prompts in addition to picture series. A closer examination of writing tasks highlighted that some studies (e.g., Shintani & Ellis, 2013; Vyatkina, 2010) allowed the use of online/electronic dictionaries during the writing process, whereas in two studies (Karim & Nassaji, 2018; Lee & Yoon, 2020) students were not allowed to use any reference materials or to discuss the pictures with other members of the group during the writing sessions. These trivial variances relating to the nature of writing tasks must have impacted outcomes.

Another variable relating to writing tasks is time constraints. Most studies (79%) were timed and 12% failed to report information about time constraints. Specifically, the length of production measures can be used as an index of L2 writing proficiency under time constraints (Wolfe-Quintero et al., 1998). The amount of writing produced within 30 minutes can indicate L2 writers' proficiency levels and the length of writing is also closely associated with the number of errors that a student may make which would impact the accuracy of written texts. For example, in Chandler's (2003) study, when calculations of error rate on the first and fifth writing assignments were made, text length was controlled by adjusting the measure of errors per 100 words, as the assignments did not yield texts of the same length. As for the length of time, we found different methods of limitations (i.e., word limit, page limit, or time limit). In all means of limitations, a wide range of differences was noted making comparability across studies impossible. Precisely, 26% of the selected studies limited the writing tasks up to 30 minutes, whereas 14% included writing tasks which lasted ten minutes during the treatment sessions and 30 minutes in pre-and post-tests.

In line with the Skills Acquisition Theory (DeKeyser, 2007), a balance between explicit instruction and extensive practice is among requisite conditions for linguistic accuracy gains. Therefore, the number of writing tasks (i.e., the amount of practice) that L2 writers are asked to complete during the intervention process is another key factor which most probably mediates the effects of WCF. The present review revealed that about a quarter

of the studies (26%) included three written tasks on three testing occasions (i.e., pretest, posttest, and delayed posttest), whereas 21% asked the participants to write four texts during the intervention period. The dynamic WCF studies (*n* = 6) (e.g., Hartshorn & Evans, 2015; Kurzer, 2018) invited participants to attempt many short texts to ensure that writing practice is extensive and manageable during the whole treatment process.

The other two critical issues in WCF concern whether participants are required to revise their writing following the feedback and whether new writing tasks are used as indicators of improved linguistic accuracy. The inclusion of the revision process ensures that their attention has been drawn to a single or multiple aspects of writing that they need to improve. Most studies (81%) reported on students' revision as a mandatory step following WCF. However, critical debates lingered with respect to the value of revision studies, as they fail to demonstrate that the effects of WCF can be carried over to new texts (see Ferris, 2010; Truscott, 2007, for detailed discussions). Due to criticism, claiming whether improvement in accuracy in students' revised texts is an indication of learning, 93% of studies included new writing tasks and compared the accuracy gains between students' initial and new texts.

Different Measures of Linguistic Accuracy in WCF Studies

The effectiveness of WCF is primarily measured by assessing the improvement of linguistic accuracy of learners' texts using outcome measures such as error ratio, obligatory occasion analysis, and holistic scoring (see Table 4). Among the diversity of accuracy measures, error ratio (Chandler, 2003) was the most frequently used in 40.4% of the 42 studies (e.g., Bonilla Lopez et. al., 2017; Mawlawi Diab, 2015; Riazantseva, 2012; Sheen et. al., 2009; Van Beuningen et. al., 2008, 2012). Riazantseva's (2012), for instance, claimed that error ratio is one of the accuracy measures used in earlier studies to correlate with measures of language proficiency in which the accuracy was measured as "a ratio of the total number of errors to the total numbers of words in the sample" (p. 425).

Depending on the scope of feedback (i.e., focused versus unfocused), the calculation of the percentage of errors varied. For example, in Mawlawi Diab's (2015) focused feedback study, the percentage of pronoun agreement errors in students' writing samples were calculated using the formula: (the number of pronoun agreement error/ the total number of words per essay) × 100, whereas in an unfocused feedback study (e.g., Nicolas-Conesa et. al., 2019), error ratio percentage for the writing tasks was calculated using the formula: (total number of errors/ total number of words) × 100. Other studies (e.g., Nicolas-Conesa et al., 2019; Truscott & Hsu, 2008) compared error rates in students' initial essays and their subsequent revisions or in their initial and final texts to examine the effect of error feedback.

Table 4. Measures of Linguistic Accuracy.

Type	Measures	No. of studies	%
1	Error ratio	17	40.4
2	Obligatory occasion analysis	14	33.3
3	Holistic ratings	4	9.5
4	Error-free clause ratio	3	7.1
5	Other measures (e.g., error counts, error reduction, error-free T-unit ratio, Jacobs scale, suppliance in obligatory content analysis etc.)	8	19.0

Another trend in the reviewed studies showed that more than 30% performed an obligatory occasion analysis to assess the accuracy gains of the targeted linguistic features (e.g., Benson & DeKeyser, 2018; Bitchener, 2008; Bitchener et. al., 2005; Ellis et. al., 2008; Sheen, 2007; Shintani & Ellis, 2013). Bitchener et al. (2005), for instance, compared the efficacy of different feedback types (i.e., direct written feedback and student-researcher 5-minute individual conference and direct written feedback only) on ESL student writing, targeting three types of errors: prepositions, past simple tense, and indefinite article. Accuracy performance was calculated as the percentage of correct usage of each targeted linguistic feature. All obligatory occasions of the target forms in each script were identified and each occasion was then inspected to determine whether they were correct or incorrect, e.g., three correct uses of the targeted linguistic form from the ten obligatory occasions mean a 30% accuracy rate. Other studies calculated the accuracy gains in a similar vein, although the targeted linguistic forms and the feedback types varied across them.

The third widely-used measure in assessing the overall writing quality of students' texts (9.5%; e.g., Chandler, 2003; Evans et. al., 2010; Vyatkina, 2010) was holistic rating. These studies made use of holistic ratings or evaluations as a secondary measurement to assess the overall writing quality of students' texts. Several studies have highlighted the importance of utilizing two accuracy measures, as assessing writing accuracy seems difficult, considering both linguistic accuracy and content. For example, Evans et al. (2010) made use of error-free clause to total clause ratio and holistic scoring to assess students' accuracy improvement. The teacher assigned a score using a holistic scoring rubric which accounted 75 percent for linguistic accuracy and 25 percent for the content. The authors concluded that similar improvement patterns were found between these accuracy measures. Along the same lines, another study (Vyatkina, 2010) investigated the extent to which different feedback types (i.e., direct, coded, and un-coded feedback) benefited students' accuracy of 15 specific error types including lexical choice, noun-related errors, and verb-related errors by comparing their error rate changes between the rough draft and the final draft of three essays. Holistic evaluations were used to assess linguistic accuracy and other dimensions of writing such as the content, relevance, creativity, and complexity.

The importance of reported reliability estimates of the accuracy measures is worth pinpointing: the lack of reliable measures can be one of the reasons contributing to

controversial findings in WCF research. Over half of the studies (59.5%) examined interrater reliability scores on the assignment of errors to the targeted categories, and only a few studies (16.6%) failed to report any reliability scores on error identification (see Table 5). The rest of the inquiries measured intra- and inter-rater reliability scores and reported high reliability estimates. Further results showed that rigorous studies (e.g., Ellis et al., 2008; Hartshorn & Evans, 2012; Sheen, 2007; Sheen et al., 2009) provided important insights into detailed scoring guidelines to assessing free-production writing tests in line with recent calls for replicable research.

Table 5. Studies Reporting Inter- and Intra-rater Reliability.

Studies with no inter-/intra-rater reliability	Ashwell 2000; Fazio 2001; Chandler 2003 (sample 2); Van Beuningen et al. 2008; Rummel and Bitchener 2015; Stefanou and Révész 2015; Kim, Choi et al. 2020
Studies reporting inter-rater reliability	Ferris and Roberts 2001; Bitchener et al. 2005; Ferris 2006; Sheen 2007; Bitchener 2008; Bitchener and Knoch 2008; Bitchener and Knoch 2009; Sheen et al. 2009; Bitchener and Knoch 2010a; Bitchener and Knoch 2010b; Evans et al. 2010; Hartshorn et al. 2010; Vyatkina 2010; Evans et al. 2011; Hartshorn and Evans 2012; Mirzaii and Aliabadi 2013; Shintani and Ellis 2013; Shintani et al. 2014; Hartshorn and Evans 2015; Mawlawi Diab 2015; Shintani and Ellis 2015; Bonilla Lopez et al. 2017; Benson and DeKeyser 2018; Kurzer 2018; Lee and Yoon 2020
Studies reporting intra-rater reliability	Ellis et al. 2008; Truscott and Hsu 2008
Studies reporting inter-and intra-rater reliability	Chandler 2003 (sample 1); Riazantseva 2012; Van Beuningen et al. 2012; Frear and Chiu 2015; Bonilla Lopez et al. 2018; Karim and Nassaji 2018; Nicolas-Conesa et al. 2019; Zhang 2021

Advantages and Pitfalls of Frequently Used Measures of Linguistic Accuracy

A key consideration in selecting the appropriate measure of linguistic accuracy depends on its discriminating power. As one of the commonly used measures of linguistic accuracy, error ratio has distinctive strengths that other measures tend to lack. For example, it can be used in studies regardless of the scope of feedback. The calculation of errors in a sample can be justified based on the targeted error types (either a limited or a broad coverage of error types). As indicated earlier, error ratio has been utilized as the accuracy measure in both focused- (e.g., Fazio, 2001; Mawlawi Diab, 2015) and unfocused-feedback studies (e.g., Ferris, 2006; Van Beuningen et al., 2012). Taking a step further in the relevance of error ratio in WCF studies, Sheen et al. (2009) examined

the extent to which focused and unfocused feedback facilitated students' accuracy by utilizing the error ratio as the accuracy measure.

A key limitation of error ratio arises from the discriminating power between the severities of errors, although it is useful for quantifying error distribution in a written sample (Riazantseva, 2012). It is unlikely to be an issue in studies targeting linguistic forms; however, it can be a problem in studies targeting a broad number of error categories. Taking the limitations into account, WCF researchers should define what is considered an error and what is not and offer detailed scoring guidelines for raters to follow. Furthermore, the average length of student sample texts should be considered in cases where text length is not controlled. Van Beuningen et al. (2012), for instance, noted that due to the relatively short texts (i.e., around 120 words), a 10-word ratio was used rather than the common 100-word ratio.

Obligatory occasion analysis is well-suited for focused feedback studies (e.g, Bitchener & Knoch, 2010a, 2010b; Shintani & Ellis, 2013), as it affords greater discriminating power than the error ratio. However, it seems unrealistic and inefficient to identify the correct and incorrect obligatory occasions of each linguistic feature in unfocused feedback studies (Bonilla Lopez et al., 2018; Karim & Nassaji, 2018). Hartshorn and Evans (2012), for instance, argued that this method "may not be possible to identify all of the obligatory occasions for every linguistic feature; nor is it appropriate for writing samples that include no obligatory occasions for a particular linguistic feature" (p. 232) without accounting for lexical errors. Further limitations were indicated by Sheen et al. (2009): they pinpointed problems in using obligatory occasion analysis for the selected grammatical features (i.e., articles, copula 'be', regular past tense, irregular past tense, and preposition), although they shared little information on what the problems were.

The other widely-employed measure, holistic scoring, tends to be practical and ecologically valid in that learners can be evaluated on such measures by writing teachers. Evans et al. (2010), for instance, used holistic scores and noted that they reflect the fullest dimensions of writing, as they consider both linguistic accuracy and the content and "it is a fairly efficient measure for a teacher who must evaluate multiple paragraphs in a timely manner" (p. 457). However, scoring is limited in that raters may find it difficult to distinguish accuracy from other global issues such as text length and content, and thus results may not be reliable.

Conclusion

This paper synthesizes findings on three key variables: feedback treatments, writing tasks, and accuracy measures which impact the efficacy of WCF in developing learners' written accuracy. It aims to indicate how these three aspects should be considered methodologically and pedagogically to guide future studies in this rich field of inquiry. Studies in WCF research have been informed by skill-acquisition theories (i.e., extensive practice and explicit rule-based instruction). Within this framework, most studies provided opportunities for learners to engage in writing tasks followed by WCF to help improve their written accuracy over time.

The 42 primary studies included in the analysis demonstrated caveats in relation to combined feedback strategies, feedback focus, and writing tasks with high/low linguistic and cognitive demands. Though direct, indirect, and metalinguistic feedback were predominant in WCF research, the use of combined feedback strategies limits our

understanding of how particular feedback strategies function. As for feedback focus, the English article system was frequently targeted in WCF, which raised the issues of how WCF functions with other linguistic features (e.g., treatable and untreatable errors). Furthermore, meaningful relationships between writing tasks of varying cognitive demands and learners' linguistic performances were also examined (e.g., Kuiken & Vedder, 2008); however, most WCF studies have failed to take them into account, which also limits their contributions to the field of L2 writing research. Concerns were raised regarding the writing accuracy measures with their distinctive strengths and weaknesses.

With these caveats in mind, we offer recommendations for future WCF research. First, studies should examine the impact of individual feedback strategies on developing L2 written accuracy. These studies may be more appealing in terms of pedagogical practices and inform writing teachers about which feedback strategies (either explicit or implicit) should be used with their students of lower or higher proficiency levels. Second, feedback focus should take into account learners' areas of difficulties and target linguistic features in specific teaching and learning contexts. Third, more investigations should examine how writing tasks with varying cognitive demands influence learners' writing performance. Such studies would contribute to the existing body of research and allow us to understand how features of writing tasks impact linguistic accuracy. As for the writing tasks used in L2 writing studies, Polio and Park (2016) claimed that most authors failed to control for topic, genre, or writing task conditions (timed/untimed, in-class/home assignments) and these variations made it difficult to determine which language changes were related to development and to task differences. Therefore, researchers should investigate how different writing genres, topics, and writing task conditions influence learners' writing performance, especially written accuracy.

Fourth, noting the affordances and limitations of available accuracy measures, studies should apply at least two accuracy measures (e.g., error ratio and holistic ratings) in a single study and investigate whether similar patterns of development are found on these measures. In addition to accuracy gains, more research is needed to consider matters of complexity and fluency to find out how attention to accuracy impacts other dimensions of language proficiency. Polio and Shea (2014) investigated the relationships between accuracy and complexity and they suggested negative associations between these two constructs. Similarly, Bruton (2010) questioned the complicated relationship between complexity and accuracy in L2 writing and concluded that "any measures of accuracy would have to be accompanied by a measure of complexity" (p. 496). Another key deliberation that received scant attention in WCF research is related to the factors considered in rating linguistic accuracy. Although the primary variable of interest in measuring linguistic accuracy is writers' ability to produce accurate texts, other secondary facets such as writing topic, writing prompts, and raters also determine some degrees of score variance (Evans et al., 2014). For example, topic familiarity and difficulty levels of writing prompts may cause fluctuations in accuracy scores of L2 texts. Based on these possible variances due to different writing topics and prompts, future research should take a combination of factors (i.e., the writer's ability level, the topic's difficulty, task complexity) into account in rating linguistic accuracy.

We hope that this review will serve researchers' needs when they design new studies in selecting criteria of writing tasks and accuracy measures to ensure comparability and reliability. It is our beliefs that future WCF studies will consider the possible impact of

feedback treatments, writing tasks, and accuracy measures on the efficacy of WCF and provide theoretical or empirical justification for their informed choices.

About the Authors

Nang Kham Thi is a Lecturer at the Department of English, University of Yangon, Myanmar. Currently, she is a Ph.D. student in Doctoral School of Education at the University of Szeged in Hungary. Her main research areas include second language writing, language assessment, and computer-assisted language learning.

Marianne Nikolov is Professor Emerita of English Applied Linguistics at the University of Pécs, Hungary. Early in her career, she taught English as a foreign language to young learners. Her research interests include early learning and teaching of modern languages, assessment of processes and outcomes in language education, individual differences, teachers' beliefs and practices, and language policy. For her full CV see her website: http://ies.btk.pte.hu/content/nikolov_marianne

To cite this article:

References

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*(3), 227–257. https://doi.org/10.1016/S1060-3743(00)00027-8

Benson, S., & DeKeyser, R. (2018). Effects of written corrective feedback and language aptitude on verb tense accuracy. *Language Teaching Research*, *23*(6), 702–726. https://doi.org/10.1177/1362168818770921

Biber, D., Nekrasova, T., & Horn, B. (2011). The effectiveness of feedback for L1-English and L2-writing development: a meta-analysis. In ETS Research Report Series. RR-11-05). Princton, NJ:EtS. https://doi.org/10.1002/j.2333-8504.2011.tb02241.x

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, *17*(2), 102–118. https://doi.org/10.1016/j.jslw.2007.11.004

Bitchener, J. (2012). A reflection on "the language learning potential" of written CF. *Journal of Second Language Writing*, *21*(4), 348–363. https://doi.org/10.1016/j.jslw.2012.09.006

Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research*, *12*(3), 409–431. https://doi.org/10.1177/1362168808089924

Bitchener, J., & Knoch, U. (2009). The value of a focused approach to written corrective feedback. *ELT Journal*, *63*(3), 204–211. https://doi.org/10.1093/elt/ccn043

Bitchener, J., & Knoch, U. (2010a). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, *19*, 207–217. https://doi.org/10.1016/j.jslw.2010.10.002

Bitchener, J., & Knoch, U. (2010b). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics*, *31*(2), 193–214. https://doi.org/10.1093/applin/amp016

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*(3), 191–205. https://doi.org/10.1016/j.jslw.2005.08.001

Bonilla Lopez, M., Van Steendam, E., & Buyse, K. (2017). Comprehensive corrective feedback on low and high proficiency writers: Examining attitudes and preferences. *International Journal of Applied Linguistics*, *168*(1), 91–128. https://doi.org/10.1075/itl.168.1.04bon

Bonilla Lopez, M., Van Steendam, E., Speelman, D., & Buyse, K. (2018). The differential effects of comprehensive feedback forms in the second language writing class. *Language Learning*, *68*(3), 813–850. https://doi.org/10.1111/lang.12295

Bruton, A. (2010). Another reply to Truscott on error correction: Improved situated designs over statistics. *System*, *38*(3), 491–498. https://doi.org/10.1016/j.system.2010.07.001

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*(3), 267–296. https://doi.org/10.1016/S1060-3743(03)00038-9

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (Second, pp. 94–112). Routledge.

Ellis, R. (2009a). A typology of written corrective feedback types. *ELT Journal, 63*(2)*,* 97-107. https://doi.org/10.1093/elt/ccn023

Ellis, R. (2009b). Corrective feedback and teacher development. *L2 Journal*, *1*(1), 2–18. https://doi.org/10.5070/l2.v1i1.9054

Ellis, R. (2009c). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*(3), 221–246. https://doi.org/10.1111/j.1473-4192.2009.00231.x

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*(3), 353–371. https://doi.org/10.1016/j.system.2008.02.001

Evans, N. W., Hartshorn, J. K., & Strong-Krause, D. (2011). The efficacy of dynamic written corrective feedback for university-matriculated ESL learners. *System*, *39*(2), 229–239. https://doi.org/10.1016/j.system.2011.04.012

Evans, N. W., Hartshorn, K. J., Cox, T. L., & Martin de Jel, T. (2014). Measuring written linguistic accuracy with weighted clause ratios: A question of validity. *Journal of Second Language Writing*, *24*(1), 33–50. https://doi.org/10.1016/j.jslw.2014.02.005

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. https://doi.org/10.1177/1362168810375367

Fazio, L. L. (2001). The effect of corrections and commentaries on the journal writing accuracy of minority- and majority-language students. *Journal of Second Language Writing*, *10*(4), 235–249. https://doi.org/10.1016/S1060-3743(01)00042-X

Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, *8*(1), 1–11.

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*(1), 49–62. https://doi.org/10.1016/j.jslw.2004.04.005

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 University Press. https://doi.org/10.1201/9781420037593.ch26

Ferris, D. R. (2003). *Error correction. In response to student writing: Implications for second-language students* (pp. 42–68). Routledge. https://doi.org/10.1201/9781420037593.ch26

Ferris, D. R. (2010). Second langauge writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition*, *32*(2), 181–201.

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, *10*(3), 161–184. https://doi.org/10.1016/S1060-3743(01)00039-X

Frear, D., & Chiu, Y. H. (2015). The effect of focused and unfocused indirect written corrective feedback on EFL learners' accuracy in new pieces of writing. *System*, *53*, 24–34. https://doi.org/10.1016/j.system.2015.06.006

Goldstein, L. M. (2004). Questions and answers about teacher written commentary and student revision: Teachers and students working together. *Journal of Second Language Writing*, *13*, 63–80. https://doi.org/10.1016/j.jslw.2004.04.006

Goldstein, L. (2006). Feedback and revision in second language writing: Contextual, teacher and student variables. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 185–205). Cambridge University Press.

Han, Y. (2017). Mediating and being mediated: Learner beliefs and learner engagement with written corrective feedback. *System*, *69*, 133–142. https://doi.org/10.1016/j.system.2017.07.003

Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*, *30*, 31–44.

Hartshorn, K. J., & Evans, N. W. (2012). The differential effects of comprehensive corrective feedback on L2 writing accuracy. *Journal of Linguistics and Language Teaching*, *3*(2), 25–50. https://doi.org/10.1515/iral.1971.9.4.319

Hartshorn, K. J., & Evans, N. W. (2015). The effects of dynamic written corrective feedback: A 30-week study. *Journal of Response to Writing*, *1*(2), 6–34.

Hartshorn, K. J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong-Krause, D., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly*, *44*(1), 84–109. https://doi.org/10.5054/tq.2010.213781

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, *39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *Modern Language Journal*, *99*(1), 1–18. https://doi.org/10.1111/modl.12189

Karim, K., & Nassaji, H. (2018). The revision and transfer effects of direct and indirect comprehensive corrective feedback on ESL students ' writing. *Language Teaching Research*, *24*(4), 519-539 . https://doi.org/10.1177/1362168818802469

Karim, K., & Nassaji, H. (2019). The effects of written corrective feedback: A critical synthesis of past and present research. *Instructed Second Language Acquisition*, *3*(1), 28–52. https://doi.org/10.1558/isla.37949

Kim, H. R., & Bowles, M. (2019). How deeply do second language learners process written corrective feedback? Insights gained from think-alouds. *TESOL Quarterly*, *53*(4), 913–938. https://doi.org/10.1002/tesq.522

Kim, Y., Choi, B., Kang, S., Kim, B., & Yun, H. (2020). Comparing the effects of direct and indirect synchronous written corrective feedback : Learning outcomes and students' perceptions. *Foreign Language Annals*, *53*(1), 176–199. https://doi.org/10.1111/flan.12443

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, *17*(1), 48–60. https://doi.org/10.1016/j.jslw.2007.08.003

Kurzer, K. (2018). Dynamic written corrective feedback in developmental multilingual writing classes. *TESOL Quarterly*, *52*(1), 5–33. https://doi.org/10.1002/tesq.366

Lee, I. (2004). Error correction in L2 secondary writing classrooms : The case of Hong Kong. *Journal of Second Language Writing*, *13*, 285–312. https://doi.org/10.1016/j.jslw.2004.08.001

Lee, I. (2005). Error correction in the L2 writing classroom: What do students think? *TESL Canada Journal*, *22*(2), 1–16.

Lee, I. (2008). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. *Journal of Second Language Writing*, *17*(2), 69–85. https://doi.org/10.1016/j.jslw.2007.10.001

Lee, J.-W., & Yoon, K.-O. (2020). Effects of written corrective feedback on the use of the English indefinite article in EFL learners' writing. *English Teaching*, *75*(2), 21–40. https://doi.org/10.15858/engtea.75.2.202006.21

Leeman, J. (2010). Feedback in L2 learning: Responding to errors during practice. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics*

and congnitive psychology (pp. 111–138). Cambridge University Press. https://doi.org/10.1017/cbo9780511667275.007

Li, S., & Vuono, A. (2019). Twenty- five years of research on oral and written corrective feedback in System. *System*, *84*, 93–109. https://doi.org/10.1016/j.system.2019.05.006

Lim, S. C., & Renandya, W. A. (2020). Efficacy of written corrective feedback in writing instruction: A meta-analysis. *TESL-EJ*, *24*(3), 1–26.

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. https://doi.org/10.1016/j.jslw.2015.08.011

Liu, Y., & Huang, J. (2020). The quality assurance of a national English writing assessment: Policy implications for quality improvement. *Studies in Educational Evaluation*, *67*(August), 100941. https://doi.org/10.1016/j.stueduc.2020.100941

Long, M. H. (1980). *Input, interaction, and second language acquisition*. University of California, Los Angeles.

Luo, Y., & Liu, Y. (2017). Comparison between peer feedback and automated feedback in college English writing: A case study. *Open Journal of Modern Linguistics*, *7*(04), 197–215. https://doi.org/10.4236/ojml.2017.74015

Mao, S. S., & Crosthwaite, P. (2019). Investigating written corrective feedback: (Mis)alignment of teachers' beliefs and practice. *Journal of Second Language Writing*, *45*(November 2018), 46–60. https://doi.org/10.1016/j.jslw.2019.05.004

Mao, Z., & Lee, I. (2020). Feedback scope in written corrective feedback: Analysis of empirical research in L2 contexts. *Assessing Writing*, *45*, 100469. https://doi.org/10.1016/j.asw.2020.100469

Mawlawi Diab, N. (2015). Effectiveness of written corrective feedback: Does type of error and type of correction matter? *Assessing Writing*, *24*, 16–34. https://doi.org/10.1016/j.asw.2015.02.001

Mirzaii, M., & Aliabadi, R. B. (2013). Direct and indirect written corrective feedback in the context of genre-based instruction on job application letter writing. *Journal of Writing Research*, *5*(2), 191–213. https://doi.org/10.17239/jowr-2013.05.02.2

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Journal of Clinical Epidemiology*, *62*(10), 1006–1012. https://doi.org/10.1016/j.jclinepi.2009.06.005

Nicolas-Conesa, F., Manchon, R. M., & Cerezo, L. (2019). The effect of unfocused direct and indirect written corrective feedback on rewritten texts and new texts: Looking into feedback for accuracy and feedback for acquisition. *The Modern Language Journal*, *103*(4), 848–873. https://doi.org/10.1111/modl.12592

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction : A research synthesis and quantitative meta-analysis. *Language Learning*, *50*(3), 417–528.

Park, E. S., Song, S., & Shin, Y. K. (2016). To what extent do learners benefit from indirect written corrective feedback? A study targeting learners of different proficiency

and heritage language status. *Language Teaching Research*, *20*(6), 678–699. https://doi.org/10.1177/1362168815609617

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing*, *21*(4), 375–389. https://doi.org/10.1016/j.jslw.2012.09.004

Polio, C., & Park, J.-H. (2016). Language development in second language writing. In R. M. Manchon & P. K. Matsuda (Eds.), *Handbook of second langauge and foreign language writing*. (pp. 287-306). De Gruyter Mouton.

Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing*, *26*, 10–27. https://doi.org/10.1016/j.jslw.2014.09.003

Polio, C., & Yoon, H. J. (2018). The reliability and validity of automated tools for examining variation in syntactic complexity across genres. *International Journal of Applied Linguistics* (United Kingdom), *28*(1), 165–188. https://doi.org/10.1111/ijal.12200

Riazantseva, A. (2012). Outcome measure of L2 writing as a mediator of the effects of corrective feedback on students' ability to write accurately. *System*, *40*(3), 421–430. https://doi.org/10.1016/j.system.2012.07.005

Robinson, P. (2011). Second language task complexity, the cognition hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: researching the cognition hypothesis of language learning and performance* (pp. 3–38). John Benjamins.

Rummel, S., & Bitchener, J. (2015). The effectiveness of written corrective feedback and the impact of Lao learners' beliefs have on uptake. *Australian Review of Applied Linguistics*, *1*, 66–84. https://doi.org/10.1075/aral.38.1.04rum

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133–164). John Benjamins.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, K. N. Chi, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010* (pp. 721–737). Singapore: National University of Singapore.

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, *41*(2), 255–283.

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*(4), 556–569. https://doi.org/10.1016/j.system.2009.09.002

Shintani, N., & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, *22*(3), 286–306. https://doi.org/10.1016/j.jslw.2013.03.011

Shintani, N., & Ellis, R. (2015). Does language analytical ability mediate the effect of written feedback on grammatical accuracy in second language writing? *System*, *49*, 110–119. https://doi.org/10.1016/j.system.2015.01.006

Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning*, *64*(1), 103–131. https://doi.org/10.1111/lang.12029

Sia, P. F. D., & Cheung, Y. L. (2017). Written corrective feedback in writing instruction: A qualitative synthesis of recent research. *Issues in Language Studies*, *6*(1), 61–80. https://doi.org/10.33736/ils.478.2017

Stefanou, C., & Révész, A. (2015). Direct written corrective feedback, learner differences, and the acquisition of second language article use for generic and specific plural reference. *Modern Language Journal*, *99*(2), 263–282. https://doi.org/10.1111/modl.12212

Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing. *Studies in Second Language Acquisition*, *32*, 303–334. https://doi.org/10.1017/S0272263109990532

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

Truscott, J. (2004). Evidence and conjecture on the effects of correction: A response to Chandler. *Journal of Second Language Writing*, *13*(4), 337–343. https://doi.org/10.1016/j.jslw.2004.05.002

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, *16*(4), 255–272. https://doi.org/10.1016/j.jslw.2007.06.003

Truscott, J. (2010). Further thoughts on Anthony Bruton's critique of the correction debate. *System*, *38*(4), 626–633. https://doi.org/10.1016/j.system.2010.10.003

Truscott, J., & Hsu, A. Y. ping. (2008). Error correction, revision, and learning. *Journal of Second Language Writing*, *17*(4), 292–305. https://doi.org/10.1016/j.jslw.2008.05.003

Van Beuningen, C., De Jong, N. H., & 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. https://doi.org/10.2143/itl.156.0.2034439

Van Beuningen, C., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, *62*(1), 1–41. https://doi.org/10.1111/j.1467-9922.2011.00674.x

Vyatkina, N. (2010). The effectiveness of written corrective feedback in teaching beginning German. *Foreign Language Annals*, *43*(4), 671–689. https://doi.org/10.1111/j.1944-9720.2010.01108.x

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. University of Hawaii Press.

Yoon, H. J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly*, *51*(2), 275–301. https://doi.org/10.1002/tesq.296

Zhang, T. (2021). The effect of highly focused versus mid-focused written corrective feedback on EFL learners' explicit and implicit knowledge development. *System*, *99*, 102493. https://doi.org/10.1016/j.system.2021.102493