Castledown

A Systematic Review of Written Corrective Feedback Research in ESL/EFL Contexts

SIN WANG CHONG [a]

[a] The Education University of Hong Kong
Email: iswchong@eduhk.hk

Abstract

In the past decade (2007–2017), research on written corrective feedback (WCF) has been proliferating in English-as-a-Second-Language (ESL) and English-as-a-Foreign-Language (EFL) contexts, and new developments in this field of research are evident. To synthesize the latest advancement in WCF research, a systematic review of recent literature on WCF was conducted to identify current research trends and provide an agenda for future WCF studies. This study was conducted following the seven stages of systematic review suggested by Petticrew and Roberts (2008). In this article, content analysis was conducted on abstracts of 41 WCF primary studies published between 1997 and 2017 in SSCI-indexed journals in the fields of TESOL, language learning, and technology and education using a text-mining tool called Leximancer. Twenty-two word-level concepts were identified, which were grouped into five themes: types of WCF, types of writing tasks, demographics of participants, research design/methods, and types of errors. Based on the systematic review, two research tasks are identified to provide an agenda for future research.

Introduction

The efficacy of written corrective feedback (WCF) has been a frequent topic of discussion and debate ever since Truscott (1996) published his seminal article in Language Learning arguing against error correction in L2 writing classes. Despite the proliferation of WCF studies in the past two decades, researchers' interest in this controversial topic have not faded but escalated in the past 10 years. A search of the websites of the leading SSCI-indexed journals in the fields of TESOL, language learning, and technology and education reveals an exponential growth in the number of primary studies published on this topic between 2007 and 2017. A total of 34 primary studies on WCF in ESL/EFL contexts were published in these journals in this period; on the contrary, there were only seven published primary studies on the same topic between 1997 and 2006.

In light of the mounting interest in WCF research in ESL/EFL contexts, a number of narrative reviews

on this topic have been published (Bitchener, 2012; Ferris, 2012; Lee, 2012). These three reviews shed important light on the following areas related to WCF research and practice:

• Bitchener (2012) summarized findings from WCF studies and argued that future WCF research should consider varying the pedagogical factors (e.g. the number of feedback treatments students are given) and examine student uptake from a sociocultural perspective. The focus of Bitchener's article is on the pedagogical implications of WCF research.
• Ferris (2012) summarized WCF studies in the form of a research timeline to provide a "historical overview" of the development of the topic (p. 446). Ferris did not focus exclusively on ESL/EFL studies but included studies conducted in foreign language contexts.
• Lee (2012) attempted to bridge the research-practice divide, focusing on reviewing WCF studies conducted in naturalistic classroom environment.

While these narrative reviews written by leading scholars in the field "provide experts' intuitive, experiential and explicit perspectives in focused topics," the methodology used in these reviews is less objective and rigorous (Pae, 2015, p. 417). For example, narrative reviews failed to "reveal how the decisions were made about relevance of studies and the validity of the included studies" (Collins & Fauser, 2005, p. 104). As such, a systematic and objective approach to content analysis of research literature, which is made possible by the development of automated content analysis (ACA) technology, namely NVivo and NAXQDA, is warranted in order to inform practice and future research through collective implications from published work (Major & Savin-Baden, 2010). More recently, text-mining tools, which are a kind of ACA technology, have been widely used to summarize and synthesize educational research, especially in areas which have witnessed rapid development (e.g., technology in education (Lee, Watson, & Watson, 2019), distance education (Zawacki-Richter & Naidu, 2016)).

As a rapidly expanding field, it is argued that WCF researchers can benefit from the use of systematic review and text-mining techniques to identify themes and gaps in the existing literature to inform their studies. The purpose of this article is twofold: (1) to review and synthesize findings from WCF research in ESL/EFL contexts published between 1997 and 2017 using a text-mining tool, Leximancer, and (2) suggest research tasks for future WCF studies based on the findings.

Background Literature

In the past decades, there has been a mushrooming of WCF studies which aim to address the question, "how effective is WCF in improving the linguistic accuracy of students' writing?" (Bitchener & Knoch, 2009, 2010; Farrokhi & Sattarpour, 2012; Sheen, 2007; Shintani & Ellis, 2013; refer to Chong, 2018, for a summary of recent studies). A great number of these studies examined the efficacy of direct and focused WCF on L2 university students' writing performance. In particular, findings from these quasi-experimental studies were able to demonstrate, through the inclusion of a pretest, posttest, and sometimes a delayed posttest, that direct and focused WCF exerts a positive influence on students' linguistic accuracy of word-level grammatical features (e.g., English articles, prepositions, the simple past tense) in writing. Moreover, comparing students' linguistic accuracy of a targeted linguistic item in the control group and treatment group(s), findings from these studies suggest that students in the treatment groups (those who received WCF) outperformed students in the control group who did not receive any feedback on their grammar performance. One sub-strand of this line of research compares the effectiveness of various kinds of WCF. To date, research has found that WCF, when given in a more explicit and corrective manner (i.e., direct WCF), is more conducive to students' uptake than more instructional WCF (e.g., metalinguistic explanation) (Shintani, Ellis, & Suzuki, 2014).

A second and emergent line of studies focus on perceptions of stakeholders of WCF (mostly from the

students' perspective). These studies answer three broad questions: (1) What is the perception of teachers and students towards WCF? (2) How do teachers go about giving WCF, and how do students go about acting on WCF? (3) What factors affect teachers' practice and students' uptake of WCF? (Ene & Kosobucki, 2016; Han, 2017; Junqueira & Payant, 2015; McMartin-Miller, 2014; Simard, Guénette, & Bergeron, 2015). Studies along this line of research suggest that ESL/EFL learners showed a preference for WCF given in a comprehensive manner, although they acknowledged that such a large amount of feedback does not always lead to successful self-corrections of errors (Ene & Kosobucki, 2016). There are many reasons which lead to students' lack of engagement with teachers' WCF; some addressed in the current literature include students' misunderstanding (Simard *et al.*, 2015), students' beliefs (Han, 2017), and students' agency (Han, 2019). Among the few studies which focus on the teachers' perspectives, they identify a number of mismatches between ESL teachers' beliefs and practices towards WCF. For instance, Lee (2003) found that Hong Kong secondary school English teachers believed that focused WCF (error correction on a number of grammatical features) would yield more pedagogical value than comprehensive WCF (error correction on all grammatical features), but most of the teachers practiced comprehensive WCF because of various external factors, including work appraisal. In the same study, findings suggested that even though the teachers devoted a lot of time giving WCF comprehensively, they were skeptical that the feedback would benefit students in the long run.

In the past few years, there have been some attempts to synthesize research findings from WCF studies. In their meta-analysis of 21 primary studies, Kang and Han (2015) investigated the effectiveness of different types of WCF on students' grammatical accuracy. Their findings demonstrated that WCF is effective in improving students' grammatical accuracy in writing, but this effect is mediated by a number of learner factors (e.g., learners' language proficiency) and contextual factors (e.g., educational setting, the writing task). In another study published in the same year, Liu and Brown (2015) conducted a methodological synthesis review on almost 50 published WCF studies and dissertations. The synthesis underscored a number of methodological constraints, including diverse WCF strategies being investigated in the same treatment group, low ecological validity, and differences in measurement of learners' linguistic accuracy. While there have been attempts to synthesize WCF studies, these two meta-analysis/synthesis articles focused primarily on one type of WCF study, i.e., quasi-experimental studies; WCF studies which adopt a sociocultural and ecological perspective were not included. In view of this, a more thorough systematic review which takes into consideration both types of WCF studies is warranted in order to map the landscape of published WCF studies to inform practice and research.

Systematic Review and Text-mining

Systematic Review

Systematic review is a "protocol-driven and quality-focused approach" to summarizing research evidence to inform research and practice (Bearman, Smith, Carbone, Slade, Baik, Hughes-Warrington, & Neumann, 2012, p. 625). As the name suggests, systematic review differs from other types of review, namely literature review, narrative review vis-à-vis its structure, comprehensiveness, and replicability. While systematic review has been commonly used to summarize "evidence-based practice" in certain educational research fields (e.g., technology in education, health education), there has been a paucity of systematic reviews in applied linguistics, language education, and higher education literature. There are four reasons for such a dearth of systematic review studies in these research fields. First, it is time-consuming to conduct systematic reviews. Researchers of systematic reviews follow strict protocols related to literature search and synthesis of findings. These protocols range from seven to nine steps. For instance, Gough (2007) suggested a nine-step process for conducting systematic reviews:

1.   Setting the research question(s)
2.   Establishing inclusion and exclusion criteria
3.   Selecting and determining the search strategies e.g., keywords, sources
4.   Screening the searched articles based on the inclusion and exclusion criteria
5.   Reporting the search strategies
6.   Deciphering relevant data from the selected studies
7.   Evaluating the methodological rigor of the selected studies
8.   Synthesizing findings from the selected studies using qualitative and/or quantitative methods
9.   Communicating these findings in an accessible manner and drawing implications from the findings

The systematic review framework, which is employed by the present study, comprises seven steps (Petticrew and Roberts, 2008):

1.   Devising research questions
2.   Specifying the types of studies
3.   Setting inclusion and exclusion criteria for literature search
4.   Conducting a literature search
5.   Screening and appraising the search results using the criteria
6.   Synthesizing findings
7.   Identifying similarities and differences in the findings

Second, systematic reviews may not be widely adopted in some educational research fields where a plethora of research approaches are used, including positivist, interpretivist, post-structuralist, and social justice (Bearman *et al.*, 2012). It is especially true in research fields where interdisciplinary research flourishes. Systematic reviews are not encouraged in research fields influenced by diverse research cultures because of the challenges they pose when attempting to synthesize findings from very different (or at times, opposing) research paradigms. Third, the term systematic review is used loosely in some research fields. Bearman *et al.* (2012) referred to a study on computer-assisted learning using the term systematic review to denote a "methodical examination of a particular set of documents" (p. 626). Elsewhere, systematic review simply means a structured way to locate publications, without taking into consideration the methodology used in synthesizing research findings (e.g., using meta-analysis or qualitative research synthesis). Finally, there is a lack of thorough understanding between the different types of reviews, namely narrative/critical review, systematic review, and scoping review. Narrative review, also called critical review, "presents a particular perspective on the literature, framed entirely through the perspective of the author" (Bearman *et al.*, 2012, p. 626). Narrative reviews, when written by experts in the topics concerned, are sometimes called "state-of-the-art" articles. On the other hand, systematic reviews are different from narrative reviews in a sense that they follow a specific protocol to conduct literature research and synthesize findings in order to generate collective research evidence to provide answers to some focused research questions. Scoping review, despite being outside the scope of the present study, is worth mentioning to provide readers with a clear distinction among the three types of review. Scoping reviews are very much akin to systematic reviews in a way that the former also utilizes "a rigorous and transparent method" (Pham, Rajī, Greig, Sargeant, Papadopoulos, & McEwen, 2014, p. 372). Nonetheless, scoping reviews differ from systematic reviews in terms of their purpose: scoping reviews aim to "map the existing literature in the field of interest in terms of the volume, nature, and characteristics of the primary research" (Pham *et al.*, 2014, p. 371) and can serve as a preliminary study to analyzing the feasibility and worthiness of conducting a systematic review.

The values of systematic reviews to educational research are manifold (Bearman *et al.*, 2012). Given

the transparency of the methodology adopted to conduct a literature search and synthesize findings, readers are provided with adequate information to evaluate the quality of the evidence as well as the methodology used for review. Moreover, employing a set of pre-determined inclusion and exclusion criteria, systematic reviews enable a relatively more objective and comprehensive selection of studies on a topic of interest, avoiding the researchers' biases to include only the best-known work. Furthermore, a well-written systematic review provides collective evidence on educational practices in a reader-friendly manner which is accessible and concise. In making research findings more accessible, it increases the likelihood that such findings will be read by policy makers and professionals in the field, which bridges the chasm between research and practice.

Text-mining

Text-mining, which usually includes three activities, namely information retrieval, information extraction, and data mining, refers to the retrieval of "information from unstructured text and to present the distilled knowledge to users in a concise form" (Thomas, McNaught, & Ananiadou, 2011, p. 2). One frequently employed method of text-mining is the use of ACA technology which, as Thomas, McNaught, and Ananiadou (2011) note, "automatically identifies and extracts terms from text" (p. 2). For the present systematic review of WCF literature, an ACA text-mining tool called Leximancer was used. Leximancer, an ACA software developed by Andrew Smith in 2000, generates themes, concepts, and ideas based on semantic and relational extractions in a bid to "make the analyst aware of the global context and significance of concepts and to help avoid fixation on particular anecdotal evidence, which may be atypical or erroneous" (Smith & Humphreys, 2006, p. 262).

Leximancer was selected as the tool because it has been increasingly used in educational research in other areas (e.g., distance education, physical education, general education) (Hyndman & Pill, 2017; Zawacki-Richter *et al.*, 2016). Additionally, Leximancer is chosen over other similar tools (e.g., Rapidminer) because of its user-friendliness. Leximancer can automatically generate a list of most frequently-appearing word-level concepts that appear in the selected documents and a concept map which shows interconnected concepts after completing a few simple steps (Figures 1 and 2). In contrast with tools focusing on manual handling of qualitative data (e.g., NVivo), Leximancer is more powerful in extracting relevant information from a large dataset because it "has the capacity to search, add, move and merge terms, as well as extract semantic (meaning) and relational information" (Sotiriadou, Brouwers, & Le, 2014, p. 3). Compared with thematic analysis by human coders, computer-aided content analysis (e.g., using Leximancer) is found to enable researchers to handle a colossal dataset without bias and increase reliability and reproducibility of the coding process and outcome (Sotiriadou *et al.*, 2014).
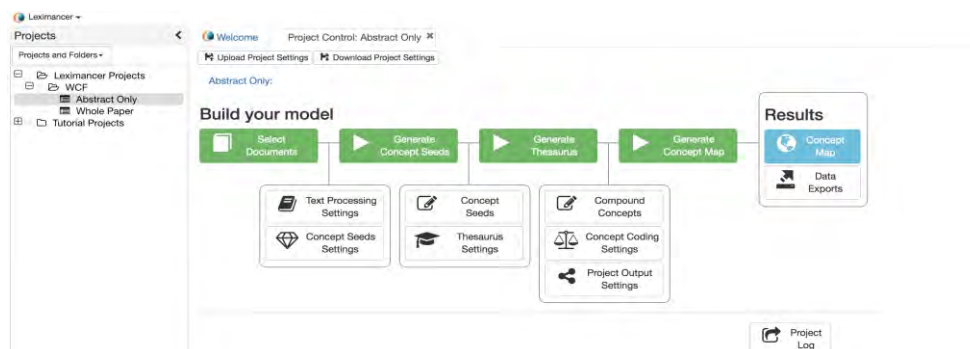

Figure 1 A five-step text-mining process in Leximancer

Figure 2 Concept map and word list generated by Leximancer

Despite being known to generate reliable outcomes, the expertise and professional judgement of the researchers when analyzing the result also play an important role because they add an additional layer of "analytical sensitivity and judgement in its interpretation" (Harwood, Gapp, & Stewart, 2015, p. 1041); therefore, it is crucial that the results be interpreted by an insider who is knowledgeable and familiar with the topic of investigation.

Method

The systematic review protocol employed by the present study includes seven steps suggested by Petticrew and Roberts (2008):

1. Devising research questions
2. Specifying the types of studies
3. Setting inclusion and exclusion criteria for the literature search
4. Conducting a literature search
5. Screening and appraising the search results using the criteria
6. Synthesizing findings
   a. Retrieving information
   b. Extracting information
   c. Mining data
   d. Interpreting data
7. Identifying similarities and differences in the findings

As for Stage 6, synthesis of findings was conducted using *Leximancer*. Specifically, findings were synthesized following the three stages of text-mining proposed by Thomas *et al.* (2011) and addressing the reminder by Harwood *et al.* (2015). To this end, a four-stage process of text-mining was used for synthesis of research findings:
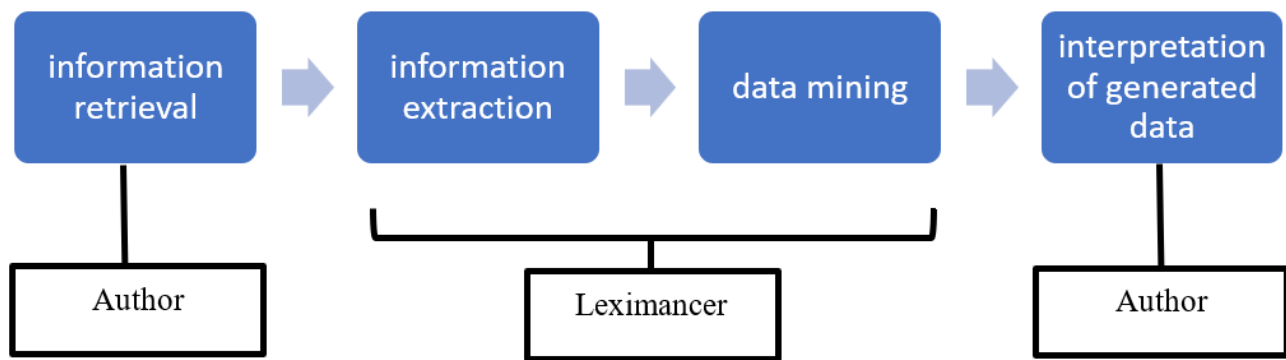
Figure 3 A four-stage process of text-mining

Devising research questions and specifying the types of studies

The present study aims at addressing the following research questions:

1. What are some WCF practices on which the included studies focused?

2. What research designs do the included studies adopt?

One of the key concerns in conducting a systematic review of literature is the setting of a timeframe (Major *et al.*, 2010). For the present review, the literature search was conducted focusing on WCF primary studies published in SSCI-indexed TESOL, language learning, and technology and education journals between 1997 and 2017. The year 1997 was chosen as the starting point of the timeframe, because Truscott's controversial article, which sparked waves of WCF research, was published in 1996. Moreover, only primary studies were included in this systematic review, meaning that conceptual papers (e.g., Storch, 2018), narrative reviews (e.g., Lee, 2012), and reporting of practice (e.g., Chong, 2017) were not included.

Table 1 Inclusion and Exclusion Criteria for Literature Search

	Include	**Exclude**
Year of publication	• 1997-2017[1]	• Before 1997
Language	• English	• Languages other than English
SSCI-indexed journal	• High impact factor in 2017 (i.e., impact factor higher than 1.00)	• Impact factor in 2017 lower than 1.00
Context of the study	• ESL or EFL	• English-as-a-first-language • Languages other than English
Focus of the study	• Corrective feedback in the written mode	• Other modes of corrective feedback e.g., oral corrective feedback • Written feedback practices that address other areas in a piece of writing, e.g., content or organization together with WCF

Setting inclusion and exclusion criteria for literature search and conducting an exhaustive literature search

Table 1 shows the inclusion and exclusion criteria for the present systematic review. The SSCI-indexed

---

[1] 2017 was selected as the last year when studies would be included because the study was conducted in 2018.

journals were identified based on two criteria: (1) The journals have a high impact factor and academic rigor in their fields; (2) the journals publish research related to WCF. Grey literature (e.g., conference proceedings) was not included because of quality control. When reviewing the articles that appeared in the search, only studies conducted in ESL/EFL contexts were retrieved. The keywords *written corrective feedback* and *error correction* were used to search for relevant research articles on the journals' websites.

Screening and appraising the search results using the criteria and synthesizing findings

Adopting the aforementioned inclusion and exclusion criteria (Table 1), a total of 41 studies were included in the present study (for a summary of the studies, refer to Appendix 1). Table 2 lists the SSCI-indexed journals and the primary studies included in the present review.

Table 2: SSCI-indexed Journals and Primary Studies Included

SSCI-indexed journal	Primary studies included
Assessing Writing	1. Lee (2003)
	2. McMartin-Miller (2014)
	3. Mawlawi Diab (2015)
	4. Ene & Kosobucki (2016)
ELT Journal	5. Bitchener & Knoch (2009a)
	6. Liao (2016)
	7. Chacón-Beltrán (2017)
Computer Assisted Language Learning	8. Shintani (2016)
Computers & Education	9. Yeh & Lo (2009)
Journal of Second Language Writing	10. Ferris & Roberts (2001)
	11. Chandler (2003)
	12. Lee (2004)
	13. Bitchener, Young, & Cameron (2005)
	14. Bitchener (2008)
	15. Truscott & Hsu (2008)
	16. Bitchener & Knoch (2010)
	17. Ferris, Liu, Sinha, & Senna (2013)
	18. Shintani & Ellis (2013)
	19. Han & Hyland (2015)
	20. Junqueira & Payant (2015)
	21. Li, Link, & Hegelheimer (2015)
Language Awareness	22. Simard, Guénette, & Bergeron (2015)
Language Learning	23. Shintani, Ellis, & Suzuki (2014)
Language Teaching Research	24. Bitchener & Knoch (2008)
	25. Zhang (2017)
The Modern Language Journal	26. Stefanou & Révész (2015)
	27. Shintani & Aubrey (2016)
System	28. Lee (1997)
	29. Gaskell & Cobb (2004)
	30. Ellis, Sheen, Murakami, & Takashima (2008)
	31. Sheen, Wright, & Moldawa (2009)
	32. Bitchener & Knoch (2009b)
	33. Evans, Hartshorn, & Krause (2011)
	34. Sampson (2012)
	35. Yeh, Lo, & Chu (2014)
	36. Frear & Chiu (2015)
	37. Shintani & Ellis (2015)
	38. Mawlawi Diab (2016)
	39. Han (2017)
TESOL Quarterly	40. Sheen (2007)
	41. Hartshorn, Evans, Merrill, Sudweeks, Strong-Krause, & Anderson (2010)

Having identified the 41 studies to be included in this systematic review, the abstract of each of the 41 studies was copied and pasted on a Word document. The Word document was then uploaded to *Leximancer* for text mining. Abstracts were included in the text mining process because "they are usually lexically dense and focus on the core concepts, themes and results of research" (Zawacki-Richter *et al.*, 2016, p. 247). A total of 22 word-level concepts were generated by *Leximancer* (Figure 4), which were grouped into five themes in response to the two research questions (Table 3). The major findings in each of the themes will be discussed in relation to the two research questions in the next section.
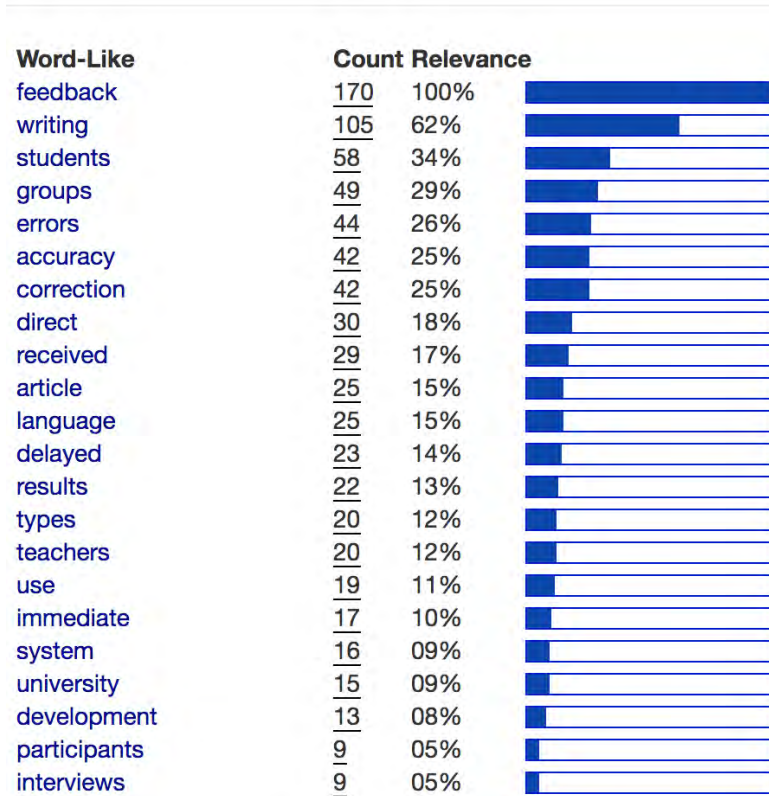
Word-Like	Count	Relevance
feedback	170	100%
writing	105	62%
students	58	34%
groups	49	29%
errors	44	26%
accuracy	42	25%
correction	42	25%
direct	30	18%
received	29	17%
article	25	15%
language	25	15%
delayed	23	14%
results	22	13%
types	20	12%
teachers	20	12%
use	19	11%
immediate	17	10%
system	16	09%
university	15	09%
development	13	08%
participants	9	05%
interviews	9	05%

Figure 4 22 word-level concepts generated by Leximancer

Table 3 Five Consolidated Themes from the Word-level Concepts

Popularity	Theme	Concept
1	Types of WCF	Feedback + correction + direct + types + system
2	Types of writing tasks	Writing
3	Participants	Students + received + teachers + university + participants
4	Research design/methods	Groups + delayed + use + immediate + development + interviews
5	Types of errors	Errors + accuracy + article + language

Findings

RQ1: What are some WCF practices on which the included studies focused?

Types of WCF

It is no surprise that the word that most frequently appears in the 41 abstracts is *feedback*, because the researched topic is on WCF. The word *correction* helps explain the function of this particular kind of written feedback, which is to correct students' errors in their written work and increase students' linguistic accuracy in writing. What is interesting is that the 41 studies investigated a broad range of WCF types, with direct WCF being the most frequently researched type of WCF. Table 4 summarizes the types of WCF covered in the reviewed studies.

Table 4 Types of WCF in the Reviewed Studies

Type of WCF	Definition	Example studies
Direct/indirect (coded) WCF	Direct WCF: Explicit correction of errors Indirect WCF: Implicit correction of errors using codes, underlining, circling	Chandler (2003) Ferris & Roberts (2001)
Focused (selective)/unfocused (comprehensive) WCF	Focused WCF: Correction of a number of pre-selected types of grammatical errors Unfocused WCF: Correction of all grammatical errors	Lee (2004) Frear & Chiu (2015)
Metalinguistic WCF (metalinguistic explanation) Synchronous/asynchronous WCF	Explanation of errors in the form of commentaries Synchronous WCF: WCF given on screen (e.g., on Google Docs) while students are writing Asynchronous WCF: WCF given on screen after students have finished writing	Bitchener (2008) Bitchener & Knoch (2010) Shintani (2016) Shintani & Aubrey (2016)
Dynamic WCF	An approach to correcting errors based on individual needs of students	Evans, Hartshorn, & Krause (2011) Hartshorn, Evans, Merrill, Sudweeks, Strong-Krause, & Anderson (2010)
Computer-generated WCF	WCF given by automatic writing evaluation (AWE) systems	Li, Link, & Hegelheimer (2015) Liao (2016)
Alternative WCF	Alternative source: WCF given by self and peers Alternative means: Concordance as WCF, WCF and rubrics	Diab (2016) Gaskell & Cobb (2004)

Two emergent types of WCF that have begun to attract the attention of researchers are computer-generated and computer-mediated WCF. The former is usually given by automated writing evaluation (AWE) systems which are either developed by service providers (e.g., *Criterion* in Li, Link, & Hegelheimer, 2015) or developed by the researchers; as for computer-mediated WCF, WCF was given by teachers using the interactive functions on self-developed online writing systems (e.g., Yeh & Lo, 2009).

Types of writing tasks

Writing is the second most frequently appearing word in the text-mining analysis. From reviewing the 41 studies, a wide range of writing tasks have been used to investigate effectiveness of WCF. These writing tasks range from sentence-level (Shintani & Aubrey, 2016), paragraph-level (Ene *et al.*, 2016; Evans, Hartshorn, & Strong-Krause, 2011), short writing, to essays. Types of short writing include picture description tasks (Bitchener, 2008; Bitchener & Knoch, 2008, 2009a, 2009b, 2010; Frear & Chiu, 2015), free-form composition (Chacón-Beltrán, 2017; Shintani, Ellis, & Suzuki, 2014), text summary (Stefanou & Révész, 2015), and Dictogloss task (Shintani & Ellis, 2015). Regarding the types of essays used in the studies, they include argumentative essay (Diab, 2015, 2016; Hartshorn *et al.*, 2010; Zhang, 2017), comparison essay (Liao, 2016), diagnostic essay (Ferris, Liu, Sinha, & Senna, 2013), five-paragraph essay (Han & Hyland, 2015), personal essay, process essay, summary essay, response essay (Li *et al.*, 2015; Yeh, Lo, Chu, 2014), and expository essay (Yeh *et al.*, 2009).

A number of writing genres were also included in the studies: narrative writing (Ellis, Sheen, Murakami, Takashima, 2008; Sampson, 2012; Sheen, 2007; Sheen, Writght, & Moldawa, 2009; Shintani & Ellis, 2013; Simard, Guénette, & Bergeron, 2015; Truscott & Hsu, 2008), business writing (Yeh, Lo, & Chu, 2014), informal letter (Bitchener, Young, Cameron, 2005), autobiographical writing (Chandler, 2003), and personal writing (Shintani, 2016).

Types of errors

The last theme that emerges from the text-mining result is the types of grammatical errors that were considered by the researchers. Among the studies which looked into focused WCF, the most commonly researched error type is about the two functional uses of the English article system, *a* and *the*. Other types of errors researched in the studies include hypothetical conditionals (e.g. Shintani *et al.*, 2015), copular *be*, the past tense, prepositions (Sheen *et al.*, 2009), weak verbs (Frea *et al.*, 2015), sentence-level errors (Gaskell *et al.*, 2004), noun ending, and wrong words (Ferris *et al.*, 2001). Judging from the above, the notions of *language* and *accuracy* in WCF studies were defined by a rather narrow perspective, dominated by word-level errors. One of the reasons why research has focused mainly on the aforesaid word-level errors is because they are rule-governed, treatable errors, in which the effectiveness of WCF is most evident (Chong, 2018; Ferris, 2011).

RQ2: What research designs do the included studies adopt?

Research designs

There were 24 studies out of the 41 reviewed that stated explicitly that a quasi-experimental design was adopted with student participants divided into one to four treatment/experimental groups (students who received different WCF treatments), usually with the inclusion of a control group (students who did not receive any WCF). These studies were interested in examining how students made use of the WCF given to improve their linguistic accuracy in future writing asks. Additionally, the quasi-experimental design of these studies includes a pretest, treatment, posttest(s). In order to measure the development of students' linguistic accuracy over a more extended period of time, more recent studies often included two posttests: an immediate posttest and a delayed posttest. Only 9 of the 41 studies included interviews as the research method/one of the research methods. These studies often adopt a case-study research design, focusing on the beliefs and perceptions of individual teachers and students towards WCF (e.g. Ferris *et al.*, 2013; Han *et al.*, 2015; Junqueira *et al.*, 2015; Lee, 2004; McMartin-Miller, 2014).

Participants

The majority of the reviewed studies included students from a range of English proficiency levels i.e. low (Bitchener, 2008), intermediate (e.g. Shintani *et al.*, 2016), and advanced (e.g. Bitchener *et al.*, 2010) as their participants. Among the studies which focused on students' uptake of WCF (how well students received and utilized WCF), most of them were conducted in colleges and universities in the U.S., New Zealand, and Taiwan, with a few which focused on secondary and high schools (e.g., Lee, 2004; Simard, *et al.*, 2015). Nevertheless, there are a dearth of studies focusing on the "giver" of WCF, the teachers (i.e., Lee, 2003; Lee, 2004; Junqueira & Payant, 2015). These studies mostly adopt a qualitative design looking into the teachers' WCF beliefs and practices.

Research Agenda

With reference to the reported findings, an agenda for WCF research is presented diagrammatically below and explained in this section.
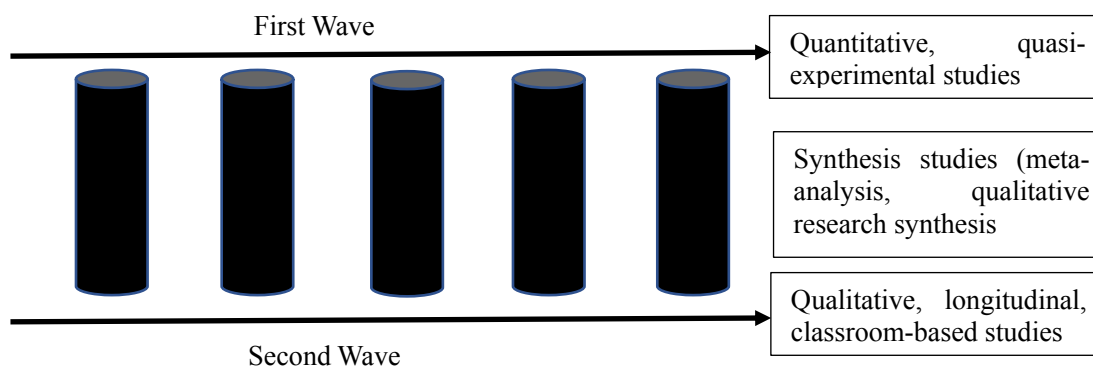

Figure 5 A diagrammatical representation of agenda for WCF research

Research task 1: Adopting a qualitative research design, future WCF studies should examine feedback practices that take place in naturalistic classroom environments across different levels of learners, especially younger ESL/EFL learners. At the same time, teachers' and students' beliefs should be tapped into to unravel the personal and contextual factors that affect teachers' WCF practices.

As reflected from the synthesized findings, the majority of the findings are dominated by quasi-experimental studies which aim to prove the efficacy of different types of WCF on the development of linguistic accuracy in students' written work. With the concerted effort of WCF researchers, an irrefutable finding is established: WCF, regardless of its type, exerts a positive effect on students' acquisition of certain linguistic features in writing. This positive effect is shown to be transferrable and long-term in some cases.

The findings from the present systematic review indicate that there is a second wave of WCF research that is being undertaken. This wave of research will be dominated by qualitative, longitudinal, classroom-based studies which focus on perceptions and practices of individual teachers and students. From an interpretivist and ecological perspective (Han, 2019), the aim of these studies is to explore and (re-)conceptualize WCF practices that are currently taking place in ESL/EFL primary, secondary, and tertiary classrooms. Taking advantage of the depth of qualitative research, various factors, namely

sociocultural, socio-emotional, and personal factors, can be explored through the use of qualitative methods, including interviews, stimulated recalls, thinking-aloud sessions, and reflective journals/blogs. To analyze such data, more sophisticated iterative and inductive coding methods, such as grounded theory, can be employed (Charmaz, 2006; Chong, 2019). The second wave of qualitative WCF studies can benefit from being informed by such theories as sociocultural theory, activity theory (Storch, 2018), and learner-context interface theory (White, Direnzo, & Bortolotto, 2016); moreover, important notions related to assessment can be incorporated to theorize the studies, including assessment literacy (Lee, 2017), feedback literacy (Carless & Boud, 2018), assessment for/as learning (Chong, 2018).

Research task 2: Synthesis studies, both qualitative (qualitative research synthesis) and quantitative (meta-analysis), should be conducted to consolidate and systemize existing knowledge on WCF, identify research gaps, and inform language policies and practices.

As shown from the synthesized findings, existing WCF studies adopt two types of research designs: quasi-experimental and naturalistic, classroom-based. Acting as "pillars" (see Figure 5) to consolidate existing findings from these two types of studies, synthesis studies on WCF research published in various forms (dissertations, conference proceedings, journal articles, research monographs) will help synthesize, consolidate and conceptualize existing knowledge garnered from these scholarly outputs. Two types of synthesis studies can be conducted: Given the domination of quasi-experimental research in the current WCF literature, *meta-analysis*, which is a statistical procedure to identify the common effect across quantitative research findings, is an appropriate type of synthesis study. Meta-analysis has been increasingly adopted in the fields of TESOL and applied linguistics (Plonsky & Brown, 2015). In preparation for the second wave of WCF research which focuses on classroom-based qualitative studies, *qualitative research synthesis* will be a positive addition to the literature because it helps to aggregate qualitative findings, which in their own standing, could hardly be transferrable. Such qualitative synthesis studies are argued to help bridge the research-practice divide, which is urged by some WCF researchers (Lee, 2013), because the condensed and crystallized research findings are easily accessed by frontline teachers. Qualitative research synthesis has already started to gain popularity in language learning research (e.g., on peer feedback by Chen, 2014, and on computer-assisted language learning by Çiftçi & Savaş, 2018). In terms of method, ACA tools such as the text-mining tool introduced in this article can be of great help to synthesists to identify prominent themes and under-explored areas for future research.

Conclusion

The present study is a systematic review of primary WCF studies in ESL and EFL contexts published between 1997 and 2017. At the outset of the paper, I present a case for systematic review by providing a succinct definition in comparison with other types of review, namely narrative review and scoping review. I then argue that systematic review is useful to (1) map the themes presented in the literature of a topic of interest in a comprehensive and objective manner, and (2) present synthesized findings in a structural fashion, guided by research questions. Adopting Petticrew and Roberts' (2008) seven-stage framework for conducting systematic review and using a text-mining tool, *Leximancer*, I present synthesized findings from 41 WCF studies published in top-tiered international refereed journals to address two research questions related to feedback practices and research designs. Regarding WCF practices, the synthesized findings indicate that there is a wide array of documented WCF strategies, most notable of which is direct WCF. At the same time, two types of WCF, computer-generated and computer-mediated WCF, have been gaining popularity. Focusing mainly on word-level grammatical

features, the findings show that WCF strategies were used in writing tasks with diverse lengths and genres. As far as the research designs of WCF studies are concerned, the synthesis results bear witness to a prevalence of quasi-experimental studies which aimed to elicit empirical evidence in support of the effectiveness of a variety of WCF strategies. Amongst those studies which adopt an interpretivist, qualitative research paradigm, the majority of the studies investigated perceptions and preferences of university students. Based on the synthesized results, I suggest two research tasks: (1) conduct WCF studies from an ecological perspective which takes into account personal and contextual factors and (2) conduct quantitative and qualitative synthesis studies to identify research gaps and inform language policies and practices.

References

Bearman, M., Smith, C. D., Carbone, A., Slade, S., Baik, C., Hughes-Warrington, M., & Neumann, D. L. (2012). Systematic review methodology in higher education. *Higher Education Research & Development, 31*(5), 625–640. https://www.doi.org/10.1080/07294360.2012.702735

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://www.doi.org/10.1016/j.jslw.2005.08.001

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

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

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

Bitchener, J., & Knoch, U. (2009b). The relative effectiveness of different types of direct written corrective feedback. *System, 37*(2), 322–329. https://www.doi.org/10.1016/j.system.2008.12.006

Bitchener, J., & Knoch, U. (2010). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing, 19*(4), 207–217. https://www.doi.org/10.1016/j.jslw.2010.10.002

Bitchener, J. (2012). Written correctivefFeedback for L2 development: Current knowledge and future research. *TESOL Quarterly, 46*(4), 855–860. https://www.doi.org/10.1002/tesq.62

Carless, D. (2016). Feedback as dialogue. *Encyclopedia of Educational Philosophy and Theory, p. 1–6*, http://link.springer.com/referenceworkentry/10.1007/978-981-287-532-7_389-1

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment and Evaluation in Higher Education, 43*(8), 1315–1325.

Chacón-Beltrán, R. (2017). Free-form writing: computerized feedback for self-correction. *ELT Journal, 71*(2), 141–149. https://www.doi.org/10.1093/elt/ccw064

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://www.doi.org/10.1016/s1060-3743(03)00038-9

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis.* Thousand Oaks, CA: SAGE Publications.

Chen, T. (2014). Technology-supported peer feedback in ESL/EFL writing classes: a research synthesis. *Computer Assisted Language Learning, 29*(2), 365–397. https://www.doi.org/10.1080/09588221.2014.960942

Chong, S. W. (2017). A three-stage model for implementing focused written corrective feedback. *TESL Canada Journal, 34*(2), 71–82.

Chong, S. W. (2018). Three paradigms of classroom assessment: Implications for written feedback research. *Language Assessment Quarterly, 15*(4), 330–347. https://www.doi.org/10.1080/15434303.2017.1405423

Chong, S. W. (2019). College students' perception of e-feedback: A grounded theory perspective. *Assessment & Evaluation in Higher Education, 44*(7), 1090–1105. https://www.doi.org/10.1080/02602938.2019.1572067

Çiftçi, E. Y., & Savaş, P. (2018). The role of telecollaboration in language and intercultural learning: A synthesis of studies published between 2010 and 2015. *ReCALL, 30*(3), 278–298. https://www.doi.org/10.1017/s0958344017000313

Collins, J. A., & Fauser, B. C. J. M. (2005). Balancing the strengths of systematic and narrative reviews. *Human Reproduction Update, 11*(2), 103–104. https://www.doi.org/10.1093/humupd/dmh058

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://www.doi.org/10.1016/j.system.2008.02.001

Ene, E., & Kosobucki, V. (2016). Rubrics and corrective feedback in ESL writing: A longitudinal case study of an L2 writer. *Assessing Writing, 30*, 3–20. https://www.doi.org/10.1016/j.asw.2016.06.003

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

Farrokhi, F., & Sattarpour, S. (2012). The Effects of Direct Written Corrective Feedback on Improvement of Grammatical Accuracy of High- proficient L2 Learners. *World Journal of Education, 2*(2), 49–57. https://www.doi.org/10.5430/wje.v2n2p49

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes. *Journal of Second Language Writing, 10*(3), 161–184. https://www.doi.org/10.1016/s1060-3743(01)00039

Ferris, D. R. (2012). Written corrective feedback in second language acquisition and writing studies. *Language Teaching, 45*(4), 446–459. https://www.doi.org/10.1017/s0261444812000250

Ferris, D. R., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing, 22*(3), 307–329. https://www.doi.org/10.1016/j.jslw.2012.09.00

Frear, D., & Chiu, Y. (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://www.doi.org/10.1016/j.system.2015.06.006

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System, 32*(3), 301–319. https://www.doi.org/10.1016/j.system.2004.04.001

Gough, D. (2007). Weight of Evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education, 22*(2), 213– 228. https://www.doi.org/10.1080/02671520701296189

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. https://www.doi.org/10.1016/j.jslw.2015.08.002

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

Han, Y. (2019). Written corrective feedback from an ecological perspective: The interaction between the context and individual learners. *System, 80*, 288–303. https://www.doi.org/10.1016/j.system.2018.12.009

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://www.doi.org/10.5054/tq.2010.213781

Harwood, I., Gapp, R. P., & Stewart, H. J. (2015). Cross-check for completeness: Exploring a novel

use of Leximancer in a grounded theory study. *The Qualitative Report, 20*, 1029–1045. Retrieved from http://nsuworks.nova.edu/tqr/

Hyndman, B., & Pill, S. (2017). What's in a concept? A Leximancer text mining analysis of physical literacy across the international literature. *European Physical Education Review, 24*(3), 292–313. https://www.doi.org/10.1177/1356336x17690312

Junqueira, L., & Payant, C. (2015). "I just want to do it right, but it's so hard": A novice teacher's written feedback beliefs and practices. *Journal of Second Language Writing, 27*, 19–36. https://www.doi.org/ 10.1016/j.jslw.2014.11.001

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. https://www.doi.org/10.1111/modl.12189

Lee, D., Watson, S. L., & Watson, W. R. (2019). Systematic literature review on self-regulated learning in massive open online courses. *Australasian Journal of Educational Technology, 35*(1), 28–41. https://www.doi.org/10.14742/ajet.3749

Lee, I. (1997). ESL learners' performance in error correction in writing: Some implications for teaching. *System, 25*(4), 465–477. https://www.doi.org/10.1016/s0346-251x(97)00045-6

Lee, I. (2003). L2 writing teachers' perspectives, practices and problems regarding error feedback. *Assessing Writing, 8*(3), 216–237. https://www.doi.org/10.1016/j.asw.2003.08.002

Lee, I. (2004). Error correction in L2 secondary writing classrooms: The case of Hong Kong. *Journal of Second Language Writing, 13*(4), 285–312. https://www.doi.org/10.1016/j.jslw.2004.08.00

Lee, I. (2012). Research into practice: Written corrective feedback. *Language Teaching, 46*(1), 108–119. https://www.doi.org/10.1017/s0261444812000390

Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Singapore: Springer.

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing, 27*, 1–18. https://www.doi.org/10.1016/j.jslw.2014.10.004

Liao, H. C. (2016). Using automated writing evaluation to reduce grammar errors in writing. *ELT Journal*, 70(3), 308–319. https://www.doi.org/10.1093/elt/ccv058

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

Major, C. H., & Savin-Baden, M. (2010). *An introduction to qualitative research synthesis: Managing the information explosion in social science research*. New York, NY: Routledge.

Major, C. H., & Savin-Baden, M. (2011). Integration of qualitative evidence: towards construction of academic knowledge in social science and professional fields. *Qualitative Research, 11*(6), 645–663. https://www.doi.org/10.1177/1468794111413367

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

Mawlawi Diab, N. (2016). A comparison of peer, teacher and self-feedback on the reduction of language errors in student essays. *System, 57*, 55–65. https://www.doi.org/10.1016/j.system.2015.12.014

McMartin-Miller, C. (2014). How much feedback is enough?: Instructor practices and student attitudes toward error treatment in second language writing. *Assessing Writing, 19*, 24–35. https://www.doi.org/10.1016/j.asw.2013.11.003

Pae, C.-U. (2015). Why systematic review rather than narrative review?. *Psychiatry Investigation, 12*(3), 417. https://www.doi.org/10.4306/pi.2015.12.3.417

Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. Malden, MA: Blackwell.

Pham, M. T., Raji , A., Greig, J. D., Sargeant, J. M., Papadopoulos, A., & McEwen, S. A. (2014). A scoping review of scoping reviews: advancing the approach and enhancing the consistency. *Research Synthesis Methods, 5*(4), 371–385. https://www.doi.org/10.1002/jrsm.1123

Plonsky, L., & Brown, D. (2014). Domain definition and search techniques in meta-analyses of L2 research (Or why 18 meta-analyses of feedback have different results). *Second Language Research, 31*(2), 267–278. https://www.doi.org/10.1177/0267658314536436

Sampson, A. (2012). Coded and uncoded error feedback: Effects on error frequencies in adult Colombian EFL learners' writing. *System, 40*(4), 494–504. https://www.doi.org/10.1016/j.system.2012.10.001

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. https://www.doi.org/10.1002/j.1545-7249.2007.tb00059.x

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://www.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://www.doi.org/10.1016/j.jslw.2013.03.011

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://www.doi.org/10.1111/lang.12029

Shintani, N. (2016). The effects of computer-mediated synchronous and asynchronous direct corrective feedback on writing: a case study. *Computer Assisted Language Learning, 29*(3), 517–538. https://www.doi.org/10.1080/09588221.2014.993400

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://www.doi.org/10.1016/j.system.2015.01.006

Shintani, N., & Aubrey, S. (2016). The effectiveness of synchronous and asynchronous written corrective feedback on grammatical accuracy in a computer-mediated environment. *The Modern Language Journal, 100*(1), 296–319. https://www.doi.org/10.1111/modl.12317

Simard, D., Guénette, D., & Bergeron, A. (2015). L2 learners' interpretation and understanding of written corrective feedback: Insights from their metalinguistic reflections. *Language Awareness, 24*(3), 233–254. https://www.doi.org/10.1080/09658416.2015.1076432

Smith, A. E., & Humphreys, M. S. (2006). Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. *Behavior Research Methods, 38*(2), 262–279. https://www.doi.org/10.3758/bf03192778

Sotiriadou, P., Brouwers, J., & Le, T.-A. (2014). Choosing a qualitative data analysis tool: a comparison of NVivo and Leximancer. *Annals of Leisure Research, 17*(2), 218–234. https://www.doi.org/10.1080/11745398.2014.902292

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. *The Modern Language Journal, 99*(2), 263–282. https://www.doi.org/10.1111/modl.12212

Storch, N. (2018). Written corrective feedback from sociocultural theoretical perspectives: A research agenda. *Language Teaching, 51*(2), 262–277. https://www.doi.org/10.1017/s0261444818000034

Thomas, D. A. (2014). Searching for significance in unstructured data: Text mining with Leximancer. *European Educational Research Journal, 13*(2), 235–256. https://www.doi.org/10.2304/eerj.2014.13.2.235

Thomas, J., McNaught, J., & Ananiadou, S. (2011). Applications of text mining within systematic

reviews. *Research Synthesis Methods, 2*(1), 1–14. https://www.doi.org/10.1002/jrsm.27

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*(2), 327–369. https://www.doi.org/10.1111/j.1467-1770.1996.tb01238.x

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

White, C., Direnzo, R., & Bortolotto, C. (2016). The learner-context interface: Emergent issues of affect and identity in technology-mediated language learning spaces. *System, 62*, 3–14. https://www.doi.org/10.1016/j.system.2016.06.008

Yeh, S.W., & Lo, J. J. (2009). Using online annotations to support error correction and corrective feedback. *Computers & Education, 52*(4), 882–892. https://www.doi.org/10.1016/j.compedu.2008.12.014

Yeh, S. W., Lo, J. J., & Chu, H. M. (2014). Application of online annotations to develop a web-based Error Correction Practice System for English writing instruction. *System, 47*, 39–52. https://www.doi.org/10.1016/j.system.2014.09.015

Zawacki-Richter, O., & Naidu, S. (2016). Mapping research trends from 35 years of publications inDistance Education. *Distance Education, 37*(3), 245–269. https://www.doi.org/10.1080/01587919.2016.1185079

Zhang, X. (2017). Reading–writing integrated tasks, comprehensive corrective feedback, and EFL writing development. *Language Teaching Research, 21*(2), 217–240. https://www.doi.org/10.1177/1362168815623291

Author biodata

Sin Wang Chong is Lecturer at the Education University of Hong Kong and a project manager/researcher at Research Institute for Learner Autonomy Education, Kanda University of International Studies, Japan. He was a guest lecturer at National Institute of Education, Nanyang Technological University, Singapore. Chong researches and has published in such areas as feedback (written corrective feedback, peer feedback, assessment feedback in higher education), language assessment (assessment of, for, as learning), language testing, second language writing, computer-assisted language learning, and learner/teacher autonomy. Chong is Associate Editor of Innovation in Language Learning and Teaching (Taylor & Francis).

Appendix 1: Overview of the 41 primary studies on WCF

SSCI-indexed journals in TESOL and applied linguistics	Primary studies on WCF included in the review	Context/participants	Type of WCF (focus, if any)	Research method/design	Major findings
Assessing Writing	Lee (2003)	ESL English secondary school teachers in Hong Kong	Detailed and comprehensive WCF and coded/indirect WCF	Teacher questionnaires, telephone interviews	Generally, teachers adopted comprehensive WCF although they did not believe in its effectiveness.
	McMartin-Miller (2014)	Instructors and undergraduates taking an ESL composition course in a U.S. university	Comprehensive and selective WCF	Interviews	• Instructors' WCF practices were flexible and different. • Students preferred comprehensive WCF.
	Mawlawi Diab (2015)	Students attending a sophomore-level ESL course at a U.S. university	Direct WCF and metalinguistic WCF (pronoun agreement and lexical errors)	Quasi-experimental design with a pretest, immediate and delayed posttests	• At the immediate posttest, a significant difference was noted in the direct metalinguistic group. • At the delayed posttest, a significant difference was noted in the direct metalinguistic group.
	Ene & Kosobucki (2016)	An ESL student attending a pre-university language study program in the U.S.	WCF and rubrics	Analysis of the student's compositions and the teacher feedback	• The use of mandatory rubrics discouraged teachers from giving personalized WCF. • The ESL student valued teachers' WCF.
ELT Journal	Bitchener & Knoch (2009)	Low-intermediate ESL students in Auckland	Focused WCF (two functional uses of the English article system i.e. 'a' and 'the')	Quasi-experimental design with a pretest, immediate posttest, and three delayed posttests	Students who received focused WCF outperformed the control group in all the posttests who received no WCF in the target language feature.
	Liao (2016)	ESL Taiwanese university students	WCF using an automated writing evaluation (AWE) system	A time-series research design with the student compositions being analyzed using descriptive analysis and paired-samples *t* tests	Generally, the number of errors in the students' compositions reduced in the revised texts and a new composition.
	Chacón-Beltrán (2017)	Spanish speaking As and B1 level (CEFR) English learners	WCF given by an AWE and corpus-based system	Analysis of student compositions	WCF focusing on an increasing number and type of errors were made available on the AWE system.

Computer Assisted Language Learning	Shintani (2016)	Two Japanese university English learners	Focused computer-mediated synchronous and asynchronous WCF	A case study to analyze the video-recorded on-screen writing process and a post-writing interview	• Both types of WCF promoted noticing-the-gap and metalinguistic understanding of the target feature. • Synchronous WCF was more effective in promoting self-correction of errors.
Computers & Education	Yeh & Lo (2009)	EFL college freshmen in Taiwan	Computer-mediated WCF given in an interactive online system	Quasi-experimental design with two experimental groups receiving different WCF treatments (paper-WCF and online-WCF)	The group of students who received WCF on the system performed better in recognizing writing errors.
Journal of Second Language Writing	Ferris & Roberts (2001)	ESL university students in a U.S. university	Focused and indirect WCF by using codes and underlining (verb errors, noun ending errors, article errors, wrong word, sentence structure)	Quasi-experimental design with two experimental groups receiving different WCF treatments and a control group receiving no WCF	• The two experimental groups which received WCF performed better in the self-editing task. • No difference was noted between the two experimental groups.
	Chandler (2003)	ESL first- and second-year students at a U.S. conservatory	Direct and indirect WCF	Study 1: Quasi-experimental design with one experimental group (who were asked to correct all errors underlined by the teacher-researcher) and one control group (who were not asked to correct the errors underlined by the teacher-researcher) Study 2: Analysis of teachers' WCF and errors made by students; student questionnaires	• Study 1: Students who were asked to correct the errors wrote more accurate compositions than those who did not. • Study 2: Direct WCF was effective in leading to students' accurate revision and students preferred direct WCF.
	Lee (2004)	ESL English secondary school teachers and students in Hong	Focused and comprehensive WCF	Teacher surveys, interviews, a teacher	Both teachers and students preferred comprehensive WCF.

		Kong		error correction task, student surveys	Only about half of the WCF provided by the teachers were accurate.
	Bitchener, Young, & Cameron (2005)	Post-intermediate adult migrant English learners in New Zealand	Focused and direct WCF (prepositions, the past tense, the definite article)	Quasi-experimental design with three experimental groups receiving different WCF treatments	WCF, together with oral feedback, was effective in increasing the accuracy of students' use of the past tense and the definite article.
	Bitchener (2008)	Low-intermediate international ESL students in Auckland	Direct WCF and metalinguistic WCF (two functional uses of the English article system i.e. 'a' and 'the')	Quasi-experimental design with three experimental groups receiving different WCF treatments and one control group receiving no WCF; a pretest-immediate-posttest-delayed-posttest design	Students who received WCF in various ways outperformed students in the control group in both the immediate and delayed posttests.
	Truscott & Hsu (2008)	EFL graduate students in Taiwan	Indirect WCF by underlining errors	Quasi-experimental design with one treatment group (who revised their compositions after receiving WCF) and one control group (who revised their compositions without WCF)	Students in the treatment group revised their compositions more successfully than those in the control group. However, this effect was only short-term.
	Bitchener & Knoch (2010)	Advanced ESL learners at a U.S. university	Metalinguistic WCF and indirect WCF by circling errors (two functional uses of the English article system i.e. 'a' and 'the')	Quasi-experimental design with three experimental groups receiving different WCF treatments and one control group receiving no WCF; a pretest-immediate-posttest-delayed-posttest design	Accuracy of the target language item was significantly different between the experimental groups and the control group in both immediate and delayed posttest.
	Ferris, Liu, Sinha, & Senna (2013)	ESL learners in a writing class at a U.S. university	Focused WCF	Student background questionnaires, student texts, interviews, field notes	Students perceived that WCF given by teachers was useful but might not always facilitate self-correction.

	Shintani & Ellis (2013)	Low-intermediate ESL students in an intensive English language program in the U.S.	Direct WCF and metalinguistic WCF (the English indefinite article)	Quasi-experimental design with two experimental groups receiving different WCF treatments and one control group who received no WCF; eye-tracking, interview	While direct WCF did not lead to an increase in accuracy in the target language feature, metalinguistic WCF was effective in doing so.
	Han & Hyland (2015)	Non-English major Chinese EFL students	Direct WCF, indirect WCF, indirect WCF with revision clues, indirect WCF with clarification requests	A case study analyzing data collected from students' compositions, interviews, retrospective verbal reports, writing conferences	Learner engagement with WCF varied among students because of individual differences.
	Junqueira & Payant (2015)	A pre-service ESL writing teacher	Direct WCF, direct WCF with explanation, indirect WCF, indirect WCF with explanation	A case study analyzing data collected from marked students' compositions, a reflective journal, interviews	There was a mismatch between the teacher's WCF belief and practice.
	Li, Link, & Hegelheimer (2015)	Writing instructors and ESL students in a U.S. university	WCF given by an AWE system	Mixed-methods design analyzing data collected from interviews and students' compositions	The AWE system led to more revisions and WCF provided by the system improved students' writing accuracy.
Language Awareness	Simard, Guénette, & Bergeron (2015)	High school ESL students in France	Direct WCF and indirect WCF	Student questionnaires	In general, students understood the WCF they received but some misunderstanding was noted.
Language Learning	Shintani, Ellis, & Suzuki (2014)	Japanese university English learners	Direct WCF and metalinguistic WCF (indefinite article and the hypothetical conditional)	Quasi-experimental design with four experimental groups receiving different WCF treatments and one control group receiving no WCF; pretest, two posttests, background questionnaire	Increased accuracy after students received WCF was noted in their use of the hypothetical conditional.

Language Teaching Research	Bitchener & Knoch (2008)	International and migrant ESL students in Auckland	Direct WCF, metalinguistic WCF (two functional uses of the English article system i.e. 'a' and 'the')	Quasi-experimental design with three experimental groups receiving different WCF treatments and a control group receiving no WCF	Students in the experimental groups outperformed the control group in terms of their level of accuracy for the target language item. Long-term effect was noted.
	Zhang (2017)	EFL English major students in China	Comprehensive WCF	Quasi-experimental design with three treatment groups and a control group receiving no WCF; pretest-posttest-delayed-posttest design	The WCF group and English-reading-English-writing group outperformed the control and Chinese-reading-English-writing groups in the posttest.
The Modern Language Journal	Stefanou & Révész (2015)	EFL high school students in Greece	Direct WCF and metalinguistic WCF (article use for specific and generic plural reference)	Quasi-experimental design with three experimental groups receiving different WCF treatments and a control group receiving no WCF; a pretest-posttest-delayed posttest design	Evidence in favor of the provision of direct WCF was yielded but benefits of incorporating metalinguistic WCF remained unclear.
	Shintani & Aubrey (2016)	Intermediate-level university students of English in Japan	Focused and direct computer-mediated synchronous and asynchronous WCF	Quasi-experimental design with two experimental groups receiving different WCF treatments and a control group receiving no WCF; a pretest-posttest-delayed posttest design	Both experimental groups showed an improved accuracy in the two posttests while the control group did not.
System	Lee (1997)	ESL university students in Hong Kong	Direct WCF and indirect/coded WCF	An error correction task	Students demonstrated a limited understanding of coded WCF and that they were better at correcting surface errors.
	Gaskell & Cobb (2004)	Lower-intermediate level EFL learners in Montreal	WCF in the form of concordance (sentence-level	Comparison between the errors made in the	Students were willing to consult concordances regarding grammatical issues

				errors)	students' pre-intervention and post-intervention writing samples	and were able to correct their errors based on concordances.
	Ellis, Sheen, Murakami, & Takashima (2008)	Japanese university learners of English	Focused and comprehensive WCF (the English definite and indefinite articles)	Quasi-experimental design with two experimental groups receiving different WCF treatments and a control group receiving no WCF; a pretest-posttest-delayed posttest design	Students in both experimental groups showed a higher level of accuracy for using the English articles than the controlled groups. Both types of WCF were equally effective.
	Sheen, Wright, & Moldawa (2009)	Adult intermediate ESL learners	Focused and unfocused/comprehensive WCF (copular 'be', regular past tense, irregular past tense, preposition)	Quasi-experimental design with three experimental groups receiving different treatments and a control group receiving no WCF; a pretest-posttest-delayed posttest design	All three experimental groups showed a gain in grammatical accuracy, with the group who received focused WCF having the highest gain scores.
	Bitchener & Knoch (2009)	Low-intermediate ESL learners in Auckland	Focused direct WCF, metalinguistic WCF (two functional uses of the English article system i.e. 'a' and 'the')	Quasi-experimental design with three experimental groups receiving different WCF treatments; a pretest-posttest-delayed posttest design	No difference was found among the accuracy of the three experimental groups.
	Evans, Hartshorn, & Strong-Krause (2011)	Undergraduate ESL students in a U.S. university	Dynamic WCF	Quasi-experimental design with one experimental group receiving dynamic WCF and one control group receiving process-writing instruction; a pretest-posttest design	While the group receiving dynamic WCF showed improvement in accuracy in writing, the group adopting a process-writing approach demonstrated a depletion in accuracy.

	Sampson (2012)	Colombian university EFL learners	Direct/uncoded and indirect /coded WCF	Comparison of errors presented in students' compositions in the two experimental groups	Both coded and uncoded WCF helped students to recognize and correct errors in their compositions while uncoded WCF might have a more long-term effect.
	Yeh, Lo, & Chu (2014)	EFL freshman in a university in Taiwan	Computer-mediated WCF given in an interactive online system	A pretest-posttest design	The system was effective in improving the accuracy of students' compositions and their performance in peer error correction.
	Frear & Chiu (2015)	EFL college students in Taiwan	Focused and unfocused/comprehensive indirect WCF (weak verbs)	Quasi-experimental design with two experimental groups receiving different WCF treatments and a control group receiving no WCF; a pretest-posttest-delayed posttest design	Both experimental groups performed better in the two posttests than the control group.
	Shintani & Ellis (2015)	Japanese university learners of English	Direct WCF and metalinguistic WCF (past hypothetical conditional, indefinite article)	A correlational study examining students' language analytical ability (LAA) mediated their writing accuracy	Learners with a stronger LAA were able to comprehend and utilize both types of WCF.
	Mawlawi Diab (2016)	EFL learners in a university in Lebanon	WCF given by teacher, peer, and self (pronoun agreement errors)	Quasi-experimental design with three experimental groups receiving different WCF treatments; a pretest-posttest-delayed posttest design	In the immediate posttest, self-feedback group significantly decreased the number of lexical errors when compared with the other two groups.
	Han (2017)	EFL university students in China	Direct, indirect/coded and comprehensive WCF	A multiple-case study analyzing data collected from interviews, retrospective verbal reports, and reflective accounts	Various learner beliefs affect students' engagement with WCF.
TESOL Quarterly	Sheen (2007)	Adult intermediate ESL learners	Focused, direct, and metalinguistic WCF	Quasi-experimental design with two	Both experimental groups performed better than the control group in the posttests. The

				experimental groups receiving different WCF treatments and a control group receiving no WCF; a pretest-posttest-delayed posttest design	group who received direct and metalinguistic WCF outperformed the group who received only direct WCF in the delayed posttests.
	Hartshorn, Evans, Merrill, Sudweeks, Strong-Krause, & Anderson (2010)	Advanced-low and advanced-mid ESL students in a university in U.S.	Dynamic WCF	Quasi-experimental design including an experimental group and a contrast group; a pretest-pottest design	Students who received dynamic WCF (the experimental group) achieved higher accuracy scores than those who were in the contrast group.