Methods of Research on Depth of Processing for Written Corrective Feedback: A Scoping Review

Yoko Fujisawa
Graduate Student, Kansai University

Atsushi Doi
Graduate Student, Kansai University

Natsuko Shintani
Kansai University

Abstract

This review employed a scoping review approach to synthesize studies published between 2015 and 2024, focusing on the depth of processing (DoP) of written corrective feedback (WCF) in second language (L2) writing. The objective was to identify trends and issues in the research methods employed in this area. The publications were extracted from Google Scholar using three keywords. After screening with coding criteria, 15 studies were identified from an initial pool of 1,850 published studies. Our review suggests that most of the selected studies were conducted in university settings and compared different types of WCF as independent variables, with DoP and accuracy as dependent variables. To measure DoP, the two primary methods adopted were think-aloud protocols (TA) and written languaging (WL). The levels of DoP varied across the studies, ranging from two to five. Several gaps in the existing DoP studies were identified and discussed, providing recommendations for future research.

Keywords: depth of processing, written corrective feedback, scoping review

1. Introduction

A scoping review is a type of literature review that aims to map out the existing research on a particular topic or field to examine the breadth and nature of studies and research in the area of interest, and to identify gaps and potential areas for future research within the existing literature (Hanzawa et al., 2024). Unlike systematic reviews, which aim to synthesize empirical evidence to answer a specific research question, scoping

reviews have a broader focus on identifying the nature, extent, and range of research within a given area. Several studies in the field of applied linguistics have employed scoping reviews to explore diverse topics, such as language learning strategies (Chong & Reinders, 2022), and task-based language teaching (Tullock & Ortega, 2017). This review employed a scoping review approach to explore the research methods in research on DoP of WCF in L2 writing.

Research on WCF has gained significant momentum, with previous studies demonstrating its effectiveness in facilitating language learning and improving accuracy in revised writing (e.g., Bitchener & Storch, 2016; Shintani et al., 2014; Van Beuningen et al., 2012). The majority of these studies have focused on the effects of different types of WCF, particularly on revising discrete linguistic forms in revised drafts or new pieces of writing. These studies have reported small to moderate beneficial effects of WCF, although many variables mediate these effects (Kang & Han, 2015).

In addition to product-oriented research, recent years have seen growing attention to process-oriented research examining how learners process WCF (e.g., Shintani & Ellis, 2015; Simard et al., 2015). Investigating WCF processing is an intriguing aspect for both researchers and educators. Theoretically, L2 learning theories, such as the Noticing Hypothesis and the Interaction Hypothesis, assume that WCF can lead to L2 learning through noticing and cognitive comparison during processing. Pedagogically, as Leow (2015) rightly points out, while WCF is an external intervention, it is crucial to elucidate the internal perspective, i.e., how learners utilize the WCF.

2. Theoretical background

Schmidt's Noticing Hypothesis (Schmidt, 1990) is one of the most influential theories in second language acquisition (SLA). To provide a definition of *noticing*, Schmidt introduces *consciousness* as a broad category of human cognition, divided into awareness, intention, and knowledge. Consciousness as awareness has three levels: perception, noticing, and understanding. Noticing refers to focal awareness where one pays attention to a stimulus as a private experience that can be verbally reported. While noticing alone does not lead to acquisition, the Noticing Hypothesis claims that learners can acquire linguistic input that they notice, or nothing is learned unless it has been noticed (Lightbown & Spada, 2013).

The effects of WCF on learners' noticing have attracted researchers, leading to extensive research on this topic (e.g., Coyle & Cánovas Guirao, 2019; Hanaoka, 2007; Hanaoka & Izumi, 2012; Izumi & Bigelow, 2000; Lázaro-Ibarrola, 2021; Lindgren & Sullivan, 2003; Qi & Lapkin, 2001). More recent studies (e.g., Suzuki, 2012, 2017) have

categorized these levels of linguistic processing during writing and WCF processing as lower (noticing) or higher (understanding). These levels correspond to Schmidt's (1990) notions of awareness at the level of noticing versus understanding. Studies classified awareness levels based on presence or absence of linguistic reasoning in TA and reflections.

Leow (2015) criticizes that noticing feedback does not guarantee deeper processing or restructuring of prior knowledge. He argues that the Noticing Hypothesis does not account for how prior knowledge interacts with noticed feedback and DoP needed to restructure interlanguage and thus, noticing is a low level of awareness. Leow (2015) proposed the concept of DoP, which refers to the cognitive effort, analysis, and elaboration that learners use when encoding input, including activating prior knowledge, testing hypotheses, and forming rules (Leow, 2015). Examining DoP in L2 writing is particularly important because factors like the slower pace of writing/ processing and visibility of written WCF might facilitate deeper processing.

Some researchers reviewed studies on DoP in L2 writing. Cerezo et al., (2019), for example, summarized seven studies that examined the levels of DoP or levels of awareness. Cerezo et al. included studies informed by three different theoretical underpinnings: four out of seven studies were based on the Noticing Hypothesis (Kuiken & Vedder, 2002; Qi & Lapkin, 2001; Sachs & Polio, 2007; Suzuki, 2012), two based on the notion of engagement (Storch, 2008; Wigglesworth & Storch, 2012), and one based on DoP (Caras, 2019). Although the review is informative for our understanding of research on DoP, there is a lack of focus on theoretical underpinnings. Considering the fact that more studies have been done on DoP in recent years and there is still no paper reviewing the methods measuring DoP among the studies, it is important to investigate updated trends in methodology for investigating DoP in L2 writing. To address this issue, the current study aims to synthesize the research methods used to analyze DoP, focusing on studies examining WCF on L2 writing. Our review focuses entirely on research based on Leow's (2015) DoP, synthesizes the research methods used in such studies, and explores the unresolved issues to offer recommendations for future research.

3. Research questions

In this review, we identify patterns in the research designs and data collection methods used in studies exploring how deeply students process different error correction techniques in L2 writing. We also aim to highlight the issues with the research methods used in these studies. Two research questions are set.

1. What are the trends in research methods used in studies investigating DoP

elicited by WCF?

2. What are the issues with the research methods used in studies investigating DoP?

4. Methods

4.1 Selection of studies

To address the research questions, we focused on published studies that measured L2 learners' DoP of WCF on their writing. The literature search was carried out by consulting an electronic database, Google Scholar. We searched for studies published from 2015, which is the year Leow (2015) was published, to February 2024, when the literature search for this review started. We combined three keywords ("depth of processing," "second language," and "writing") and used them when searching literature on the database. Initially, through this database search, a total of 1,850 reports were identified.

The retrieved studies were screened based on the following four criteria to determine which to include in this review.

1. Only experimental studies that examined learners' DoP while they received WCF on L2 writing tasks (e.g., essays and blog entries) and predefined operational definitions of DoP were included. Studies employing reformulation and model text were also included if these tools were used to provide WCF on L2 writing.
2. Studies that examined cognitive processing of WCF based on the Noticing Hypothesis were excluded (e.g., Suzuki, 2017).
3. Studies that did not focus on L2 learning were excluded. For example, DeRobles (2019) was excluded because all participants were heritage language learners, not L2 learners, who had grown up in bilingual households either since birth or before the age of six.
4. In cases where the same study was published in separate papers, the paper with the most detailed description of research methods was included, while the other was excluded. For example, a dissertation (Caras, 2017) was included, while a book chapter (Caras, 2019) was excluded.

As a result, a total of 15 studies remained which satisfied all of the criteria. The studies comprised seven journal articles, five book chapters, three dissertations.

4.2 Coding criteria and procedures

The retrieved literature was first coded for the study identifier (i.e., authors and year), and publication type (journal article, book chapter, or dissertation). Furthermore,

eight features were also coded: type of institution, number of participants, independent variables, dependent variables, control group, writing task, type of WCF, data for DoP, levels of DoP (presented in Table 1).

Type of institution represented the type of institution from which the participant's data was collected. It was coded as either university, secondary (i.e., senior high school and junior high school), primary (i.e., primary school), online L2 course, and not mentioned (i.e., no information provided). Number of participants showed the total number of participants in each study.

For the writing task category (i.e., the type of writing tasks employed), six subcategories were identified: essay, decision-making task, blog entry, story retelling task (i.e., reading a story and retelling it without consulting the original text), story based on scenario (i.e., writing a story according to a provided written scenario), and story based on a picture prompt (i.e., recounting a story with a picture prompt depicting the story). Essay refers to a writing task in which participants were instructed to produce a piece of writing about a prompt with a topic. Picture-based and problem-solving tasks, which were originally developed by Gilabert (2007), were categorized into decision-making task.

We classified type of WCF (i.e., the types of feedback used in treatment) into five categories: direct WCF, metalinguistic WCF, indirect WCF, reformulation, and computer-generated WCF. The coding was conducted according to what was described in each study. For instance, "indirect corrective feedback" in Park and Kim (2019) was coded as indirect WCF and "metalinguistic error coding" in Bowles and Gastañaga (2022) was coded as metalinguistic WCF. However, there were three exceptions. A type of WCF named "underlining" (i.e., underlining errors) in Bowles and Gastañaga (2022) was coded as indirect WCF against the following explanation of indirect WCF "the teacher indicates that an error exists but does not provide the correction" (Ellis, 2009, p. 98). "Teacher feedback" in Kaivanpanah et al. (2020) was coded as metalinguistic WCF because it involved indicating what kinds of errors students had made (e.g., grammar, wrong word, and spelling) by using codes. This was in line with the following explanation of metalinguistic WCF: "The teacher provides some kind of metalinguistic clue as to the nature of the error" (Ellis, 2009, p. 98). "Indirect WCF with metalinguistic explanations" in Ma (2020) was also coded as metalinguistic WCF because it pointed out error types and offered metalinguistic explanations of how to correct the errors (but did not provide explicit corrections). Direct WCF, metalinguistic WCF, indirect WCF, and computer-generated WCF were further coded as either focused or unfocused.

The independent variables in the included studies were categorized based on the

interventions conducted in the experiments. Eight types of independent variables were identified: (1) type of WCF (i.e., comparing the impacts of two or more types of WCF on dependent variables); (2) teacher vs. computer WCF (i.e., comparing teacher-written WCF with computer-generated WCF); (3) +/– TA (TA vs. silent conditions); (4) TA, WL, and TA+WL; (5) TA in L1 vs. L2 (i.e., comparing L1 with L2); (6) individual vs. collaborative writing (i.e., comparing individual writing with collaborative one); (7) pen-and-paper vs. computer-mediated environment; and (8) proficiency levels (elementary vs. intermediate).

The dependent variables of each study were first identified using the statements in the research questions. As this review focused on DoP, the terms, which were mentioned in the RQs, "process" (Caras, 2017; Ma, 2020), "cognitively process" (Shao & Liu, 2022), and "levels of processing" (Adrada-Rafael & Filgueras-Gómez, 2019; Park & Kim, 2019) were also coded as DoP. Bowles and Gastañaga (2022) worded accuracy as "accurate error revision," and Manchón et al. (2020) worded "linguistic accuracy of the rewritten texts," while Ma (2020) worded "accuracy with le and classifiers." Therefore, each study was scrutinized, and those that measured the accuracy of the rewrite were coded as revision accuracy, while those that measured the accuracy regarding grammatical items were coded as accuracy of target items.

Additionally, two wording in the RQs from two studies (McBride, 2023; Shao & Liu, 2022) were modified when coded. McBride's (2023) research question, "How does writing in a traditional pen-and-paper environment versus writing in a computer-mediated environment affect L2 written production in terms of CAF measures?" was coded as CAF in writing. Similarly, a research question in Shao and Liu (2022), "To what extent does WCF foster the acquisition of regular past tense by L2 learners?" was coded as Grammar acquisition.

Data for DoP represented data collection instruments of DoP, which were coded as either TA or WL. Although McBride (2023) and McBride and Manchón (2023) both employed three types of data instruments: TA, WL, and TA + WL, only TA was coded. It was because the data of WL, TA + WL were analyzed based on other theoretical approaches (i.e., the Noticing Hypothesis and engagement). Levels of DoP represented how many levels of DoP were set in each study. Control group was coded either as yes or no to indicate the presence and absence of a control group in each study.

The first and second authors completed the coding. One author was responsible for half of the 15 studies, and the other one conducted the remaining half. We then checked each other's coding results and resolved all disparities or confusion through discussion.

5. Results and discussion

To answer Research Question 1 (the trends in research methods used in studies investigating DoP), we summarized the research methods of the 15 studies included in this review as shown in Table 1. In the following sections, we examine the results and seek for the answer to Research Question 2, that asked the issues with the research methods used in studies investigating DoP of WCF.

5.1 General research methods in DoP studies

The literature search targeted publications from 2015 to 2024, but as shown in Table 1, all studies were from 2019 or later, except for Caras (2017), suggesting a recent increase in interest in DoP. These included seven journal articles, five book chapters, and three dissertations. Three out of five book chapters (Adrada-Rafael & Filgueras-Gómez, 2019; Cerezo et al., 2019; Park & Kim, 2019) were published in a single handbook, *The Routledge Handbook of Second Language Research in Classroom Learning* (Leow, 2019).

Table 1

Included Studies

Study	Pub. type	Institution	*n*	Independent variables	Dependent variables	Control group	Writing task	Type of WCF	Data for DoP
Abdi Tabari et al. (2023)	Journal	University	80	Type of WCF, +/– TA	Revision quality, DoP	No	Essay	Direct, indirect WCF (all unfocused)	TA
Adrada-Rafael and Filgueras-Gómez (2019)	Book chapter	University	29	TA in L1 vs. L2	Revision accuracy, DoP	No	Story based on a picture prompt	Reformulation	TA
Bowles and Gastañaga (2022)	Journal	University	35	Type of WCF	Revision accuracy, DoP	No	Essay	Direct, metalinguistic, indirect WCF (all unfocused)	TA
Caras (2017)	Dissertation	University	61	Type of WCF	Accuracy of target items, DoP	Yes	Blog entry	Direct, metalinguistic, Indirect WCF (all unfocused)	TA
Cerezo et al. (2019)	Book chapter	University	46	Type of WCF	Revision accuracy, error noticing, DoP	Yes	Essay	Direct, indirect WCF (all unfocused)	WL
Kaivanpanah et al. (2020)	Journal	Not mentioned	60	Teacher vs. computer WCF +/– TA	Quality of writing, DoP	No	Essay	Metalinguistic, computer-generated WCF (all unfocused)	TA
Kim and Bowles (2019)	Journal	University	22	Type of WCF	DoP, error types	No	Essay	Direct WCF (unfocused), reformulation	TA
Leow et al. (2022)	Journal	University	10	Type of WCF	Subsequent compositions, curricular tests, final exam, targeted linguistic feature, over time, DoP	No	Essay	Direct, metalinguistic WCF (all unfocused)	TA
Ma (2020)	Dissertation	Online L2 course	38	Type of WCF	Accuracy of target items, engagement, DoP	No	Story based on scenario	Direct, metalinguistic WCF (all unfocused)	TA
Manchón et al. (2020)	Book chapter	University	118	Individual vs. collaborative writing	Revision accuracy, error detection, DoP	Yes	Decision-making task	Direct WCF (unfocused)	WL
McBride and Manchón (2023)	Book chapter	University	18	TA, WL, and TA+WL	Engagement, Affordance for investigating DoP	No	Decision-making task	Direct WCF (unfocused)	TA
McBride (2023)	Dissertation	University	36	TA, WL, and TA+WL Pen-and-paper vs. computer-mediated environment	CAF in writing, DoP	No	Decision-making task	Direct WCF (unfocused)	TA
Nicolás-Conesa et al. (2019)	Journal	University	46	Type of WCF	Revision accuracy, error types	Yes	Essay	Direct, indirect WCF (unfocused)	WL
Park and Kim (2019)	Book chapter	University	24	Proficiency levels (elementary vs. intermediate)	Revision accuracy, DoP	No	Essay	Indirect WCF (unfocused)	TA
Shao and Liu (2022)	Journal	Secondary	99	Type of WCF	Grammar acquisition, DoP	Yes	Story retelling task	Direct, metalinguistic, indirect WCF (all focused)	TA

As Table 1 illustrates, the majority of the included DoP studies were conducted at universities, with three exceptions. Ma (2020) conducted research in an online Mandarin Chinese course designed specifically for professionals who would live in China. The mean ages of her two experimental groups were 38 and 41, respectively. Shao and Liu (2022) investigated secondary school students. Kaivanpanah et al. (2020) did not provide information regarding the institutional setting. The imbalance in research settings being skewed towards universities has been pointed out in many previous meta-analyses. The studies on DoP reviewed here exhibit this same tendency of being predominantly conducted in university settings. Since DoP is a cognitive process, it is presumed to be influenced by the learner's cognitive levels, calling for research investigating young learners and comparative research on DoP between children and adults.

The number of participants ranged from 10 to 118. Some studies employed a within-group design, in which a single group experienced two different treatments, and their performance in the two treatments was compared. For example, in Kim and Bowles (2019), all 20 participants completed two different essay writing tasks, receiving either direct WCF or reformulation, and the researchers compared the learners' DoP when receiving different types of WCF. Similarly, Bowles and Gastañaga (2022) had 35 participants who wrote three different essays, receiving direct WCF, metalinguistic error coding, and underlining, respectively. Although in all these studies, the essay tasks were counterbalanced to control for task difficulty, the small sample sizes may limit the generalizability of the findings to the broader population.

Other studies had multiple experimental groups. Manchón et al. (2020) had 118 participants divided into two experimental groups, an individual and a collaborative writing condition, as well as a control group. They examined differences in error detection, accuracy of the rewritten writing, and DoP between the conditions. Caras (2017) had three experimental groups: unfocused direct WCF (n = 15), unfocused metalinguistic WCF (n = 16), and unfocused indirect WCF (n = 15), as well as a control group (n = 15), and examined the learners' process of WCF.

Regarding writing tasks, eight studies employed essay writing with three variations: argumentative (Kaivanpanah et al., 2020; Kim & Bowles, 2019), narrative (Ma, 2020; Nicolás-Conesa et al., 2019; Tabari et al., 2023), and others (Bowles & Gastañaga, 2022; Cerezo et al., 2019; Park & Kim, 2019). The other three studies (Manchón et al., 2020; McBride, 2023; McBride & Manchón, 2023) employed a decision-making task in which the participants explained a solution to rescue as many people as possible based on a picture of a building on fire. Adrada-Rafael and Filgueras-

Gómez (2019) asked participants to write a story based on a picture prompt. Ma (2020) had them write a story based on a scenario. Caras (2017) adopted a blog entry task in which participants wrote a blog entry on a Spanish-language site and received contextual instructions and a checklist to address specific questions. Shao and Liu (2022) adopted a story-retelling task in which the participants read a story about a past event and reproduced it in their writing without referencing the original story.

5.2 Independent variables

Nine out of 15 studies compared the effectiveness of two or more types of WCF, while the remaining six employed one type of WCF to observe DoP when receiving WCF. Among the nine comparative studies, six compared two types, direct and indirect WCF, while three compared three types, including metalinguistic WCF.

Twelve studies included direct WCF in an independent variable. While most of them only employed direct corrections on the erroneous forms, Abdi Tabari et al. (2023) provided explicit correction with metalinguistic WCF. Seven studies included indirect WCF, which was provided in various forms: highlighting (e.g., Abdi Tabari et al., 2023), underlining (e.g., Caras, 2017), and indicated by a circle (e.g., Shao & Liu, 2022). Two of the seven studies, Cerezo et al. (2019) and Nicolás-Conesa et al. (2019), provided metalinguistic codes (i.e., codes representing error category) as a part of indirect WCF. Six studies included metalinguistic WCF, five indicated errors by using error coding. Among them, four studies (Caras, 2017; Leow et al., 2022; Ma, 2020; Shao & Liu, 2022) further complemented error codes with metalinguistic explanations or samples of rules. The remained study, Shao and Liu (2022) provided a brief explanation of the target linguistic rule with direct error corrections.

Two studies (Adrada-Rafael & Filgueras-Gómez, 2019; Kim & Bowles, 2019) employed reformulation, which refers to a native-speaker rewriting students' original compositions to make them sounds native-like with bearing in mind to keep the students' ideas much as possible (Ellis, 2009). While Adrada-Rafael and Filgueras-Gómez (2019) provided the participants only with reformulation, Kim and Bowles (2019) compared it with direct WCF. Only one study, Kaivanpanah et al. (2020), employed computer-generated WCF which compared it with teacher-written metalinguistic WCF.

The overview of type of WCF suggests that measuring DoP of WCF was attempted in relatively limited situations: when learners received direct, indirect, and metalinguistic WCF. Research on DoP when given computer-generated WCF or when comparing reformulated writing or model texts with learners' own writing is limited.

All studies except one employed unfocused WCF. Shao and Liu (2022) was the

only study that used focused WCF. This is assumingly because the researchers were interested in the learners' DoP, which might differ according to targeted linguistic features. Caras (2017), for example, examined the levels of DoP for the feedback targeting different grammatical features (i.e., two target dichotomies, which were ser versus estar and the preterit versus imperfect past tense aspects) and suggested that the levels of DoP differ according to the linguistic features.

Besides impacts of WCF on dependent variables, other interventions such as writing environment (i.e., pen-and-paper vs. computer-mediated environment), how a task is completed (i.e., individual vs. collaborative writing), and proficiency levels (elementary vs. intermediate) were investigated, even though there was only one study for each of the three aspects.

10 out of the 15 studies did not include a control group. As two of them did not investigate the effects of WCF on the improvement of writing and focused entirely on DoP (Kim & Bowles, 2019) or engagement and affordances for inspecting DoP (McBride & Manchón, 2023), they did not need to include a control group. However, the other studies examined the improvements as a result of providing WCF by measuring revision accuracy and grammar acquisition. In such research, including a control group would allow stronger implications (Ellis et al., 2019). Five studies had a control group. In four of these studies, the control groups completed the writing tasks without receiving any WCF (Cerezo et al., 2019; Manchón et al., 2020; Nicolás-Conesa et al., 2019; Shao & Liu, 2022), and in the other study, the control group received commentary on the overall organization and content of their writing (Caras, 2017).

5.3 Dependent variables

For a dependent variable, eight studies measured accuracy. We identified two types of accuracy measures in the studies: six studies investigated revision accuracy, and two assessed the accuracy of specific target items. The former studies include Nicolás-Conesa et al. (2019), which quantified errors within each task and group to assess accuracy. Park and Kim (2019) employed three categories: accurate correction, inaccurate correction, and no correction to investigate participants' self-corrections. Conversely, the latter two studies evaluated accuracy on specific target items. Caras (2017) concentrated on copular verbs, gender agreement in adjectives, and past tense aspect in Spanish, while Ma (2020) focused on le and classifiers, two fundamental structures frequently used in Chinese. Only two studies explored the development of L2 proficiency beyond evaluating the accuracy of revisions. These studies assessed the complexity, accuracy, and fluency of learners' new writings (McBride, 2023) or utilized

evaluative tests administered by the educational institution where the study was conducted (Leow et al., 2022) to ascertain whether writing tasks with WCF contributed to enhancements in L2 proficiency.

All the studies included in this review examined DoP as a dependent variable, except for Nicolás-Conesa et al. (2019), which measured DoP but did not specify how the data was utilized. Concerning the measurement of DoP, 10 out of 15 studies categorized DoP into three levels: high, medium (intermediate), and low, following Leow's (2015) framework. Two studies (Kim & Bowles, 2019; Leow et al., 2022) modified Leow's (2015) coding scheme; Kim and Bowles (2019) classified DoP into two levels: high or low, while Leow et al. (2022) employed four levels of DoP: High DoP + Understanding, High DoP – Understanding, Medium DoP + Reporting, and Low DoP + Noticing. The remaining three studies (Cerezo et al., 2019; Manchón et al., 2020; Nicolás-Conesa et al., 2019) classified DoP into five levels. As previously noted, the categorization of DoP varied across studies.

Regarding the data collection method for identifying DoP, 12 out of the 15 studies employed TA, while the other three used WL. These two data collection instruments allow researchers to obtain concurrent introspective data, which refers to real-time self-reflective information of ongoing experiences. Unlike retrospective data, which is gathered after the fact and often relies on memory or existing records to reconstruct past events (e.g., interviews and stimulated recall), concurrent introspective data is more likely to provide real-time insights into cognitive processes (Sasaki, 2013). However, such concurrent introspective data collection has potential issues of reactivity and veridicality; the TA or WL themselves might interfere with the learner's natural behavior or serve as an intervention for the learners (Zhang & Zhang, 2019). In response to these concerns, Abdi Tabari et al. (2023), in the current review, treated TA as an intervention to facilitate learning and made it one of the independent variables in their study. The results suggested that the data collection mode (TA, WL) impacts learners' cognitive processes and contributes to learning. Taking further steps, McBride and Manchón (2023) involved three experimental groups (i.e., TA, WL, and TA+WL) to examine how learners' engagement and DoP differ in each data collection mode. These studies suggest that researchers should be mindful of the potential influence of TA or WL on data of learners' cognitive processes when employing these data collection methods. TA language (i.e., whether L1 or L2) was also expected to have impacts on learners' DoP in one study (Adrada-Rafael & Filgueras-Gómez, 2019) in this review, however, no other study investigated this aspect.

The methods employed for TA practice tasks varied across studies. Four studies

utilized a math calculation task (Adrada-Rafael & Filgueras-Gómez, 2019; Caras, 2017; Leow et al., 2022; Park & Kim, 2019), while three studies (Bowles & Gastañaga, 2022; Kim & Bowles, 2019; Tabari et al., 2023) incorporated a practice task similar to the actual task in which participants engaged in TA while receiving WCF. Considering the actual task participants were expected to undertake during TA, the more closely the practice task resembles the actual task, the more effectively participants may approach the latter. This aligns with Bowles's (2010) observation that although arithmetic tasks have the advantage of being non-verbal and thus not interfering with the practice itself, they may be challenging to apply in an actual task. Instead, he suggested that verbal warm-up tasks that resemble the actual task enable participants to verbalize more readily while engaging in the actual task. The remaining studies, such as Kaivanpanah et al. (2020), Ma (2020), McBride (2023), and McBride and Manchón (2023), provided instructions on TA without practice, while Shao and Liu (2022) employed a reading comprehension task as a TA practice alongside instructions.

WL has also been used in previous research to inspect cognitive activity while processing WCF (e.g., Suzuki, 2012, 2017) and was mentioned as a good option for collecting data in a real classroom without disrupting the class dynamics (Nicolás-Conesa et al., 2019). In this review, three studies (i.e., Cerezo et al., 2019; Manchón et al., 2020; Nicolás-Conesa et al., 2019) employed it as a data collection instrument of DoP. In their WL phase, the three studies asked participants to correct errors or reflect on errors already provided on their compositions, and to provide a linguistic explanation for each error and corresponding correction in a form. Two of the three studies, Cerezo et al. (2019) and Nicolás-Conesa et al. (2019) provided the following prompt (adopted from Suzuki, 2012); "Have a look at your essay and identify your errors. Why is each linguistic form incorrect? Explain it." Languages used in WL activities also varied among studies. Manchón et al. (2020) instructed them to do it in L2 (English) but permitted them to use Spanish (most of the participants' L1) if necessary, and Cerezo et al. (2019) also employed participants' L2 (i.e., English). However, no information of language in WL was provided in Nicolás-Conesa et al. (2019).

6. Conclusion

In this scoping review, we focused on the research methods employed in studies examining DoP. The review provided an overview of the existing studies investigating learners' DoP when receiving WCF. It highlighted some research gaps in the existing literature, such as targeting learners other than university students, computer-mediated writing contexts, or writing for specific purposes. The majority of the studies included

in this review employed commonly used comparisons in the field (i.e., direct, indirect, and metalinguistic WCF) and lacked exploration of other types, such as reformulation and computer-generated WCF. Research using WL is still scarce.

This review also identified some limitations in the research methods used in the DoP experiments, such as reliance on within-group designs, small sample sizes, and a lack of control groups. Another issue highlighted in this review is the potential influence of TA when examining DoP. As Shao and Liu (2022) argue, TA may serve not only as a means of obtaining introspective data but also as a potential educational intervention.

As this is a preliminary review, we acknowledge that there are more aspects to analyze in the literature to fulfill the purpose of the scoping review. The most important aspect to review is the way the existing literature differentiates the levels of DoP. As shown in the current review, studies vary in terms of how many levels they set when examining learners' DoP. The conceptual underpinnings of the DoP levels can be explored by conducting an in-depth analysis of the DoP categories in individual studies. Such analysis would provide a conceptual scope of the notion of DoP. We intend to further pursue such analysis.

While limited, we believe this review demonstrated the usefulness of scoping reviews in providing an overview of the existing literature, highlighting potential gaps, and future research directions.

References

Adrada-Rafael, S., & Filgueras-Gómez, M. (2019). Reactivity, language of think-aloud protocol, and depth of processing in the processing of reformulated feedback. In R. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 199–211). Routledge.

Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 development* (Vol. 96). Multilingual Matters.

Bowles, M. A. (2010). *The think-aloud controversy in second language research*. Routledge. https://www.routledge.com/The-Think-Aloud-Controversy-in-Second-Language-Research/Bowles/p/book/9780415994842

Bowles, M. A., & Gastañaga, K. (2022). Heritage, second and third language learner processing of written corrective feedback: Evidence from think-alouds. *Studies in Second Language Learning and Teaching, 12*(4), 675–696. https://doi.org/10.14746/ssllt.2022.12.4.7

Caras A. (2019). Written corrective feedback in compositions and the role of depth of processing. In R. Leow (Ed.), *The Routledge handbook of second language research*

in classroom learning (pp. 186–198). Routledge.

Caras, A. M. (2017). *"Why are these underlined?" Depth of processing and type of written corrective feedback in L2 Spanish compositions* (Publication No. 10601613) [Doctoral dissertation, Georgetown University]. ProQuest Dissertations and Theses Global.

Cerezo, L., Manchón, R. M., & Nicolás-Conesa, F. (2019). What do learners notice while processing written corrective feedback? A look at depth of processing via written languaging. In R. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 171–185). Routledge.

Chong, S. W., & Reinders, H. (2022). Autonomy of English language learners: A scoping review of research and practice. *Language Teaching Research, 0*(0). https://doi.org/10.1177/13621688221075812

Coyle, Y., & Cánovas Guirao, J. (2019). Learning to write in a second language: The role of guided interaction in promoting children's noticing from model texts. CLIL *Journal of Innovation and Research in Plurilingual and Pluricultural Education, 2*(1), 21. https://doi.org/10.5565/rev/clil.22

DeRobles, G. (2019). *The effects of type of written corrective feedback and level of proficiency on processing and accuracy in heritage language learners of Spanish* (Publication No. 13899079) [Doctoral dissertation, Georgetown University]. ProQuest Dissertations and Theses Global.

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

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2019). *Task-Based Language Teaching: Theory and Practice*. Cambridge: Cambridge University Press.

Hanaoka, O. (2007). Output, noticing, and learning: an investigation into the role of spontaneous attention to form in a four-stage writing task. *Language Teaching Research, 11*(4), 459–479. https://doi.org/10.1177/1362168807080963

Hanaoka, O., & Izumi, S. (2012). Noticing and uptake: Addressing pre-articulated covert problems in L2 writing. *Journal of Second Language Writing, 21*(4), 332–347. https://doi.org/10.1016/j.jslw.2012.09.008

Hanzawa, K., Suzuki, Y., Yanagisawa, A., & Fukuta, J. (2024, March 15–19). *A scoping review on task repetition studies focusing on speaking ability*. [Paper presentation]. American Accusation for Applied Linguistic Conference, Houston, Texas, United States. https://www.xcdsystem.com/aaal/program/SqDM5Li/index.cfm

Izumi, S., & Bigelow, M. (2000). Does output promote noticing and second language acquisition? *TESOL Quarterly, 34*(2), 239. https://doi.org/10.2307/3587952

Kaivanpanah, S., Alavi, M., & Meschi, R. (2020). L2 writers' processing of teacher vs. computer-generated feedback. *Journal of English Language Teaching and Learning, 12*(26), 175–215. https://doi.org/10.22034/elt.2020.11472

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

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

Lázaro-Ibarrola A., Hidalgo M. Á. (2021). Give me a second chance: Task repetition and collaborative writing with child EFL learners. *Language Teaching for Young Learners, 3*, 275–299. https://doi.org/10.1075/ltyl.20009.laz

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. Routledge.

Leow, R. P., Thinglum, A., & Leow, S. A. (2022). WCF processing in the L2 curriculum: a look at type of WCF, type of linguistic item, and L2 performance. *Studies in Second Language Learning and Teaching, 12*(4), 651–673. https://doi.org/10.14746/ssllt.2022.12.4.6

Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th ed.). Oxford University Press.

Lindgren, E., & Sullivan, K. P. H. (2003). Stimulated recall as a trigger for increasing noticing and language awareness in the L2 writing classroom: A case study of two young female writers. *Language Awareness, 12*(3–4), 172–186. https://doi.org/10.1080/09658410308667075

Ma, X. (2020). *Writing in a task-based individualized curriculum: Effectiveness of direct and indirect written corrective feedback* (Publication No. 28089503) [Doctoral dissertation, Georgetown University]. ProQuest Dissertations and Theses Global.

Manchón, R. M., Nicolás-Conesa, F., Cerezo, L., & Criado, R. (2020). L2 writers' processing of written corrective feedback. In R. M. Manchón (Ed.), *Languaging in language learning and teaching: A collection of empirical studies* (pp. 241–263). John Benjamins Publishing Company.

McBride, S. (2023). *The effects of written corrective feedback processing conditions on L2 learners' written texts and levels of depth of processing* [Doctoral dissertation, University of Murcia]. https://doi.org/10.13140/RG.2.2.10891.46881

McBride, S., & Manchón, R. M. (2023). Analysing L2 writers' processing of written corrective feedback via written languaging and think-aloud protocols. In. R. M.

Manchón & J. R. de Larios (Eds.), *Research methods in the study of L2 writing processes* (Vol. 5, pp. 337–514). John Benjamins Publishing Company.

Nicolás-Conesa, F., Manchón, 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

Park, E. S., & Kim, O. Y. (2019). Learners' engagement with indirect written corrective feedback: Depth of processing and self-correction. In R. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 212–226). Routledge.

Qi, D. S., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing, 10*(4), 277–303. https://doi.org/10.1016/S1060-3743(01)00046-7

Sasaki, M. (2013). Introspective Methods. In A. J. Kunnan (Ed.), *The Companion to Language Assessment, First Edition.* (pp. 1340–1357). Wiley.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Shao, J., & Liu, Y. (2022). Written corrective feedback, learner-internal cognitive processes, and the acquisition of regular past tense by Chinese L2 learners of English. *Applied Linguistics Review, 13*(6), 1005–1028. https://doi.org/10.1515/applirev-2019-0131

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

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://doi.org/10.1080/09658416.2015.1076432

Suzuki, W. (2012). Written languaging, direct correction, and second language writing revision. *Language Learning, 62*(4), 1110–1133. https://doi.org/10.1111/j.1467-9922.2012.00720.x

Suzuki, W. (2017). The effect of quality of written languaging on second language learning. *Writing and Pedagogy, 8*(3), 461–482. https://doi.org/10.1558/wap.27291

Tabari, M. A., Sato, M., & Wang, Y. (2023). Engagement with written corrective feedback: Examination of feedback types and think-aloud protocol as pedagogical interventions. *Language Teaching Research, 0*(0). https://doi.org/10.1177/13621688231202574

Tullock, B., & Ortega, L. (2017). Fluency and multilingualism in study abroad: Lessons from a scoping review. *System*, *71*, 7–21. https://doi.org/10.1016/j.system.2017.09.019

Van Beuningen, C. G., 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

Zhang, L. J., & Zhang, D. (2019). Think-aloud protocols. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 302–312). Routledge.