International Journal of Linguistics and Translation Studies
Volume 6, Issue 1, 2025
Homepage: http://ijlts.org/index.php/ijlts/index
DOI: https://doi.org/10.36892/ijlts.v6i1.542

Writing Development in the L2: A Look into Depth of Processing Using Verbal Protocols

Sergio Adrada-Rafael[1]
Department of Modern Languages and Literatures, Fairfield University, Fairfield, USA
sadradarafael@fairfield.edu

Rose Meschi
Sultan Qaboos University, Muscat, Oman
r.meschi@squ.edu.om

How to cite:
Adrada-Rafael, S. & Meschi, R. (2025). Writing Development in the L2: A Look into Depth of Processing Using Verbal Protocols. *International Journal of Linguistics and Translation Studies 6*(1).41-59.
https://doi.org/10.36892/ijlts.v6i1.542

ARTICLE HISTORY
Received: *19/11/2024*
Accepted: 18/01/2025

Keywords:
Depth of processing, online processing, L2 writing, verbal protocols, feedback

Abstract

Depth of processing (DOP) refers to the way information is processed in the mind and how it affects recall. Studies have examined DOP in different areas of L2, such as reading, and more recently attention has been paid to learners' writing, but it remains to be more thoroughly studied. Therefore, to better understand the role of DOP in the incorporation of corrective feedback on the writings of English as a Foreign Language (EFL) language learners, and to find out how DOP is related to changes in the learners' final written product, the present study asked 30 intermediate EFL learners having Persian as their L1 to edit their writings based on the feedback they received, either thinking aloud or being silent. Findings suggested that learners processed lexical items at a deeper level than grammatical items. A significant difference was found when comparing the quality of the two writing versions. Thinking aloud did not impact learners' performance when compared to the silent condition. These findings contribute to the strand of recent studies that have looked into the role of DOP in L2 writing and highlight how DOP contributes to L2 learners' writing development.

1. INTRODUCTION

In the field of second language acquisition (SLA), several studies have examined the mental processes of second or foreign language (L2) learners when performing a task to gain deeper insights into the internal processes of learning, and how these impact learners' performance (e.g., (Adrada-Rafael, 2017; Craik & Lockhart, 1972; Craik & Tulving, 1975; Leow & Mercer, 2015; Qi & Lapkin, 2001). Thus, concepts such as working memory, attention, awareness, noticing and levels of processing have attracted the attention of researchers. Among these, Craik and Lockhart (1972) introduced the concept of levels of processing in the context of first language (L1) learning, emphasizing that memory retention depends not only on practice, study, or attention at the moment of input but also on how new information is processed during task engagement.

[1] Corresponding author

Looking further into the processing of lexical items, Craik and Lockhart (1972) argued that processing information occurs along a continuum, ranging in two main levels, from perceptual or shallow processing to semantic, conceptual or deep processing. Hallow processing involves focusing on surface-level features, such as orthographic or phonological features, while deep processing requires greater cognitive effort and engagement with the meaning of the words, including connections to prior knowledge. Shallow processing involves rehearsing and repeating the received information as well as paying attention to the appearance of words or their pronunciation, which aids in keeping the received information in short-term memory and does not result in robust retention (Leow & Mercer, 2015). Deep processing, on the other hand, occurs when learners analyze words for meaning or their relationships to other words, resulting in stronger memory traces, enhanced retention, and greater recall (Adrada-Rafael, 2017; Leow & Mercer, 2015).

Although the effects of DOP have been extensively investigated in relation to reading tasks, their impact on writing tasks remains underexplored (e.g., Leow, 2019; Leow, Hsieh and Moreno, 2008; Morgan-Short, Heil, et al., 2012; VanPatten, 1990). To address this gap, the present study aims to contribute to the growing body of DOP literature by examining how DOP manifests in writing tasks, particularly after learners receive feedback on grammatical and lexical errors. The following sections provide a comprehensive review of key studies investigating DOP to date.

2. REVIEW OF THE LITERATURE

Craik and Lockhart's (1972) foundational framework on levels of processing (LOP) posits that a stimulus that causes deeper processing can lead to higher levels of retention, long-term memory traces and hence future retrieval. Expanding on this, Craik and Tulving (1975) conducted a series of experiments designed to manipulate DOP through different tasks. Participants were presented with lists of words and promoted to engage with the stimuli at varying levels of depth. For the shallow level, questions regarding the word's typeface were asked (e.g. if the word HOUSE is written in uppercase letters). For the medium level of processing, questions were concerned with recognizing rhyme, (e.g. does the word "house" rhyme with the word "pencil"?). Finally, for the deep level of processing, questions were concerned with the semantic features of the word, (e.g. does the word "house" fit into this sentence: "The _____ has a window"?). Results indicated that when answers were compatible, target words were better and easier recalled than the incompatible ones, highlighting the critical role of processing in memory retention. Figure 1 depicts a summary of the three processes designed in the two mentioned studies.

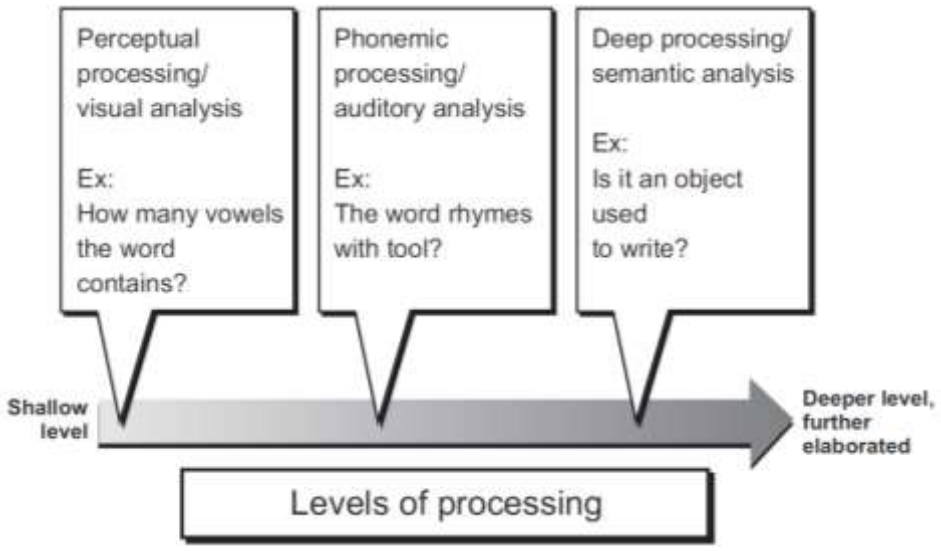


Figure 1. Diagram of levels of processing while encoding the written word "pen" (the subject can analyze it at various levels of processing) from Ekuni et al. 2011

Alt text for Figure 1 (20 words): Timeline showing three levels of processing, moving from a shallow to a deeper level, and the characteristics of each level.

The Levels of processing theory suggests that if the learner uses higher and greater cognitive efforts at the time of processing information while drawing from the prior background knowledge, then the possibility of recalling the processed information is enhanced. When a learner pays attention to input, a trace of memory is formed. The depth at which this trace has been processed and whether it will be further processed determines if it will be recalled later or it will perish easily from memory (Leow & Mercer, 2015).

2.1.DOP in SLA studies

In the field of SLA, DOP has gained attention as a framework for understanding how learners process input. In an early study, VanPatten (1990) posited that paying attention to form and grammar diverts attention from meaning and, therefore, it would potentially hinder comprehension. To test his hypothesis, he conducted a study with 202 Spanish learners from different levels who were tasked with listening to a 275-word text in Spanish. Participants were divided into three experimental groups and one control group. In the experimental groups, the first group had to focus on the lexical item *inflación,* the second on the article *la,* and the third on the third-person plural verbal morpheme *–n,* whereas in the control group locused on content. Results revealed that focusing on the article la or the third-person plural verbal morpheme –n negatively impacted comprehension, while attention to lexical items yielded similar results to the content-focused control group. These findings suggest that attention to grammar hinders processing for meaning and that attention to semantics and lexical items requires a deeper level of processing, which can be hindered if the focus is on form.

Leow, Hsieh and Moreno (2008) modified VanPatten's (1990) study in several ways. They used concurrent think aloud protocols to determine whether attention to form or meaning could affect performance. They also changed the presentation from aural to written input and turned the free recall test into a multiple-choice test. Participants were 72 English-speaking Spanish students who were randomly assigned to one of the five groups. The concurrent or online think aloud data was recorded for each participant in each condition. The results of a multiple-choice test measuring their comprehension, revealed no significant difference among the five conditions. The think-aloud data was analyzed to yield a better picture of the cognitive processes in the students' minds. Three levels of processing were reported, the deeper one for the lexical item *sol* (73%), for *la* (45%), and for *lo,* or –n (31%). The results of the think-aloud data did not establish a relationship between processing and comprehension.

Morgan-Short, Heil, Botero-Moriarty, and Ebert (2012) replicated Leow et al.'s study to further explore their findings. They added a non-think aloud group to investigate for reactivity during reading for meaning and form and increased the sample size to compensate for the limitation Leow et al. faced in generalizing the findings. While noticing a significant effect for reactivity, they reported a positive correlation between depth of processing and the comprehension score; i.e. when the processing occurred at a deeper level, the score on the comprehension test was higher indicating better understanding and hence better learning.

Depth of processing has mainly been studied using a reading task; however, in an early study Qi and Lapkin (2001) investigated depth of processing in L2 writing. In their case study with two Mandarin students, who studied English as a second language with different proficiency levels, students were instructed to complete a writing task with a picture prompt in three stages. In the first stage, they were asked to write a story based on the prompt; they received syntactic and morphological feedback in addition to focusing on stylistic and logical sequencing

problems at the discourse level. In the second stage, they received a reformulated version of their writings along with their first drafts. They had to compare the two pieces of writing while thinking aloud and their stream of thought was recorded for further investigation. Finally, they wrote a new, revised version of their essays using the modifications they had received three days in advance in stage two. After analyzing the recorded think-aloud data, it was found that deeper processing led to more corrections and higher quality writings.

In a more recent study, Adrada-Rafael (2017) investigated how 88 intermediate-level learners of Spanish acquired the Imperfect Subjunctive in Spanish under different types of instruction, which differed in their degree of explicitness: More explicit, Less explicit, and Baseline. The design followed a pre-, post-, delayed test design, and participants were presented with a reading task where the Subjunctive forms were embedded. The author addressed DOP by asking participants to think aloud during the reading phase. The study findings showed that deeper processing correlated with a higher accuracy in the production of the target form and with a greater comprehension of the reading passage. Results also revealed that participants in the More explicit condition produced more instances of processing, which could be explained by the fact that they were exposed to metalinguistic information during the instructional phase.

2.2.DOP in SLA studies with a focus on writing and feedback

Adrada-Rafael and Filgueras-Gómez (2019) partially extended Sachs and Polio's (2007) second experiment, maintaining the same three-day sequence design while introducing two groups: thinking aloud (TA) and silent. The study involved 44 advanced Spanish learners (L1 English, L2 Spanish), with 29 assigned to the TA group and 15 to the silent group. Within the TA group, 13 participants verbalized in their L1, while 16 verbalized in their L2. In the first session, they wrote a short story based on a picture prompt. During the second session, they received a reformulated version of their first drafts and were randomly assigned to one of two experimental conditions: reformulation feedback + TA in L1 or reformulation feedback + TA in L2. . In session 3, a weekend later, students re-wrote their original stories without access to the feedback. The researchers categorized the instances of processing as low, intermediate and deep according to Leow's (2015) descriptors for grammatical and lexical items. Results revealed no traces of reactivity. Additionally, more instances of processing occurred in the L2 rather than in the L1. While both grammatical and lexical items were processed at intermediate and deep levels, learners processed more deeply when thinking aloud in their L1.

Kim and Bowles (2019) investigated DOP in academic writing tasks, focusing on two types of feedback: direct correction and reformulation. 22 adult learners enrolled in an academic writing course were asked to write two argumentative essays, receiving direct correction on one and reformulation feedback on the other in a counterbalanced design. Think-aloud protocols were employed to gain insight into how participants processed these two types of feedback. The findings revealed that participants processed sentential and paragraph-level errors at a deeper level, however, they overlooked the surface-level errors with reformulation feedback. In addition, reformulations evoked a higher depth of processing for non-surface errors. However, the researchers did not find reformulation as the superior type of feedback in comparison with direct corrections, as their study lacked measures for learning outcomes and was limited to advanced learners, making the results difficult to generalize. Similar to Adrada-Rafael and Filgueras-Gómez (2019), Kim and Bowles found no significant differences between the writing performance of the think aloud and silent groups. They concluded that there might not be a one-size-fits-all kind of feedback for improving students' writing and that different errors require different types of feedback.

Especially important for the purposes of the present manuscript is Leow's model of Instructed Second Language Acquisition (ISLA) (2015), where he integrated the concept of DOP within the stages of language acquisition. In this model, he operationalized three different levels of processing (low, medium, and high) for lexical and grammatical items, including a series of

descriptors that help identify the DOP level when analyzing the TA protocols. Additionally, levels of awareness of noticing, reporting and understanding were included.

Quite recently, a series of studies have tackled the theoretical underpinnings of the writing process when being provided corrective feedback, while also accounting for DOP. Leow (2020) presented an overview of the underpinnings underlying the writing-to-learn process, offering theoretical, practical and pedagogical implications for the L2 classroom, and also differentiating it from other processes, such as writing-for-accuracy. Cerezo, Manchón and Nicolás-Conesa (2019) and Manchón, Nicolás-Conesa, Cerezo & Criado (2020) both looked into whether levels of DOP were impacted by different writing conditions (individual vs. collaborative writing) either receiving or not receiving feedback, in several dimensions of accuracy. For the latter study, participants were 118 intermediate learners of English as a foreign language at a public Spanish university. Findings revealed that the availability of feedback (or lack of it), rather than the writing condition participants were exposed to, was the key factor affecting DOP and accuracy. The authors concluded that writing with the availability of feedback led to deeper levels of processing, and, in consequence, to a greater languaging activity. Probing deeper into DOP and its effects on SLA, Leow, Thinglum & Leow (2022) conducted a longitudinal study with ten learners of Spanish at the beginner level where they examined writing performance throughout a semester, by examining different types of linguistic items (syntactic vs. morphological) with two different types of corrective feedback (direct vs. metalinguistic). Participants had to complete three compositions with their subsequent revisions. The specific addressed items were the Spanish agreement between noun (morphological) and adjective and the *gustar* structure (syntactic). Verbal protocols to account for participants' DOP were collected during the revision phase. Results showed a higher DOP when being provided metalinguistic feedback. This DOP remained high during the rewrite stage regardless of the linguistic item analyzed. The authors suggested that, when trying to understand what leads to L2 development, the type of feedback received by L2 learners in their writing might not be as important as how the said feedback is processed. Furthermore, McBride and Manchón (2023) discussed the methodological considerations to bear in mind when addressing DOP in L2 writing studies, detailing the methodological advantages of employing verbal protocols to gain a better understanding of learners' thoughts involved in the writing process. Given the findings of the empirical studies reviewed above, it can be argued that DOP is beneficial and facilitative in the analysis of L2 grammatical and lexical items.

Previous studies seem to indicate that the deeper the processing, the higher the possibility for learning, robust retention and activation of prior knowledge; therefore, resulting in a better performance. Despite the growing interest in DOP in SLA, most studies have focused on reading tasks or have examined writing tasks with a limited range of feedback types (e.g., Kim & Bowles, 2019; Leow et al., 2022; Manchón et al., 2020). Research examining the relationship between DOP and indirect coded feedback—a commonly used feedback type in classroom settings—remains scarce. Additionally, no studies, to the authors' knowledge, have explored DOP in L1 Persian-speaking learners of English. This population presents a unique opportunity to expand our understanding of how linguistic factors may influence DOP in writing tasks. Thus, the present study seeks to address these gaps by investigating how L1 Persian-speaking learners of English as an L2 process lexical and grammatical items after receiving indirect-coded feedback on a writing task and whether DOP relates to linguistic gains. Specifically, the study addresses the following research questions:

1. Do EFL language learners process lexical and grammatical items at the same level of processing after receiving teacher feedback on a writing task?

2. Does DOP relate to improvements in the final written product of L2 learners when compared to learners in a silent condition?

3. METHOD

3.1.Participants

Participants were 30 undergraduate Persian-speaking intermediate EFL learners enrolled on an EFL course at an Iranian university. Their ages ranged from 18 to 22 years old, and they had an average of 5.5 years of study of English as their L2. This group was composed of 16 females and 14 males. They all consented to participate in the study and received extra credit in exchange for their participation.

3.2.Materials

In order to make sure that all participants were homogenous in terms of their general English proficiency prior to the study, they took the Oxford Placement test (OPT). The OPT includes two sections: listening and grammar. There are 100 items in the listening section, which take approximately ten minutes to complete. The grammar section includes 100 items, which take up to 50 minutes to be completed.

Participants' writings were scored based on the ESL Composition Profile (see Appendix B) developed by Jacobs, Zinkgraf, Wormuth, Hartfiel, and Hughey (1981) which considers content, organization, vocabulary, language use, and mechanics; each one has four rating levels ranging from very poor, poor to fair, average to good, and very good to excellent. The proficiency level of the writer is described through a series of components with clear definers, as well as a numerical scale. The content component is defined on a scale ranging from 13-30, organization 7-20, vocabulary 7-20, language 5-25, and mechanics 2-5. The rating level of excellent to very good for content has a maximum score of 30 to a minimum of 27 representing a piece of writing which is "knowledgeable, substantive, (has) thorough development of thesis, relevant to assigned topic", whereas very poor content of an essay had a maximum score of 16 and a minimum score of 13 showing that the essay "does not show knowledge of subject, (is) not substantive, not pertinent, or not enough to evaluate" (Jacobs et al., 1981, p.4).

3.3.Procedure

Before they participated in the study, all participants showed a similar level of English proficiency upon completing the OPT, intermediate, which discarded that any potential differences in their results could be due to significant differences in pre-study proficiency. The study was conducted in two stages: In the first stage, participants were instructed to write an argumentative essay based on the following prompt: "*It is true that whilst many people use their public parks, this space could be used for other purposes, such as building houses for young people or developing a business area to create jobs. Explain to what extent you agree or disagree with this statement in 250 words*." They were given 40 minutes to write the essay and were not allowed to use a dictionary or any other sources to make sure that they relied solely on their own knowledge. They were asked to proofread their essays before handing them to the teacher to make sure that the possible existing errors were not mistakes or slips of pen, which could have been possibly caused by their lack of concentration and could be self-corrected by the author (James, 1998; Qi & Lapkin, 2001; Poulisse, 1999).

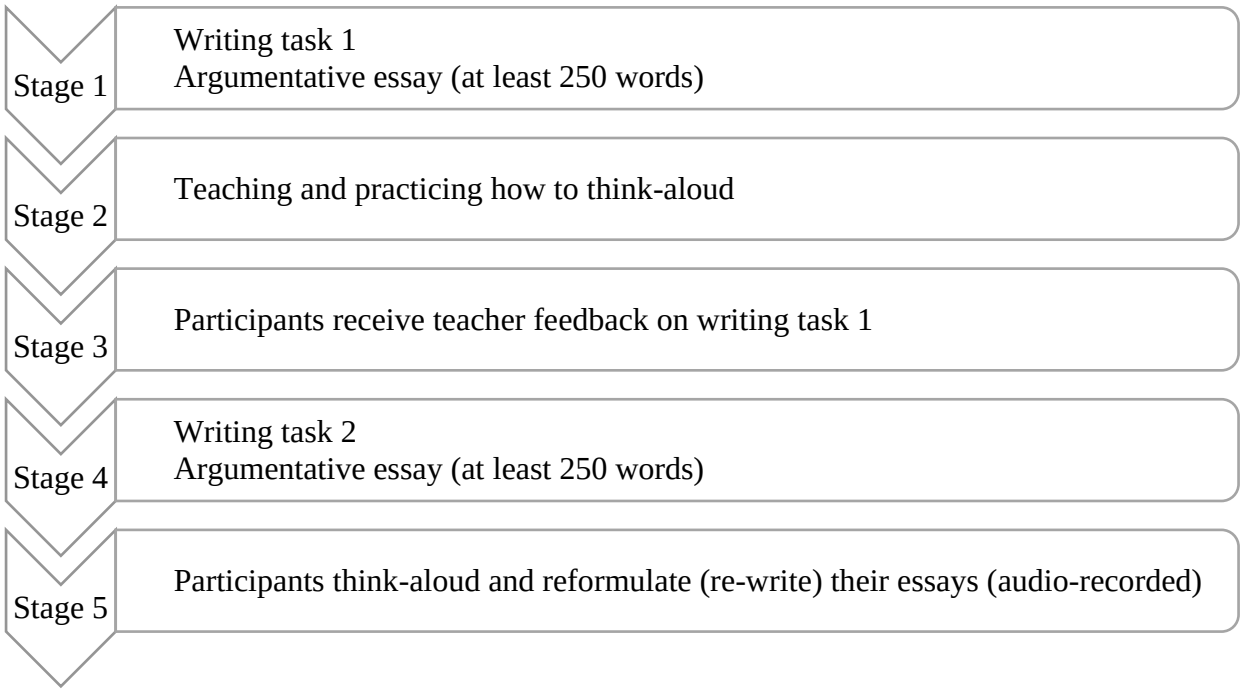


Figure 2 Procedures

Alt text for Figure 2 (24 words): Timeline showing the five stages of the study procedure, with the initial writing, the feedback received, the rewrite, and also when participants think aloud.

The collected essays were then corrected by the instructor and the researcher based on their form (grammatical accuracy), content, semantics and word choice using the ESL composition profile. Both the instructor and researcher corrected all of the essays. Very few discrepancies arose when deciding how to categorize the errors, and an agreement was reached on those differences found. Interrater reliability was calculated resulting in a high level of agreement (Cronbach alpha: .87). Students received indirect coded feedback on their essays which enabled them to produce significant gains in their writings and helped them develop metalinguistic awareness (Ferris, 2002, Shizuka, 2000; Lalande, 1982; Kahraman, 2013). According to Kim and Bowles (2019), who used verbal protocols to examine how learners process reformulation feedback and direct correction, learners tend to overlook surface-level errors and process paragraph-level errors deeply when receiving reformulation, whereas the reverse results were received for direct correction.

A week later, in the second stage of the study, participants were requested to re-write their essays based on the feedback they had received in the classroom[2]. At this stage, participants in one of the groups were asked to think aloud concurrently while revising their essays individually in a language laboratory. Prior to thinking aloud, and in order to make sure that they were familiar with the procedure, they received practice instruction on how to think aloud using a narrated PowerPoint presented in a .mov format, which was already used in Adrada-Rafael (2017). This practice aimed at helping learners describe and recognize their mistakes, which were indirectly pointed out by the teacher as well as responding and showing reaction to their mistakes. In other words, it aimed to help students explain what they were thinking while correcting their essays and hence help the researchers gain insights into their minds and

[2] The present study was part of a larger study where both teacher and computer feedback were provided, with a total of 4 groups (+/-TA). For the present manuscript, only teacher feedback with TA and silent groups is addressed.

thoughts. The researcher and two other teachers also modelled how to think aloud while reflecting on the feedback on a piece of writing. Participants could think aloud in their L1 to avoid any potential confounds on their conclusions regarding reactivity that could be caused if instructed to use their L2[3] (Adrada-Rafael & Filgueras-Gómez, 2019; Sachs & Polio, 2007). Allowing participants to think aloud in their L1 was due to the fact that nearly all learners at the intermediate level of language proficiency tend to think in their mother tongue while trying to discover the mistakes in their essays (Qi & Lapkin, 2001). They were instructed to let their thoughts flow loudly and to express what they thought. Their verbal protocols were recorded and transcribed by the researchers to understand how participants noticed their mistakes and how they analyzed their mistakes using the provided correct version.

4. RESULTS

4.1.Research Question 1:

In order to examine the role of DOP with the incorporation of corrective feedback on the writings of EFL language learners, each participant's level of processing for lexical and grammatical items was determined by following Leow's operationalization of DOP (2015) (See Appendix C). Table 1 displays the frequency analysis for each level of processing (low, intermediate and high) regarding lexical or grammatical items before and after receiving teacher feedback.

Table 1. *Descriptive Statistics of lexical and grammatical items (Teacher feedback-Think aloud)*

Participants (Teacher feedback – T/A)	Lexical items			Grammatical items			Total Low	Total Medium	Total High	Accuracy		Quality	
	Low	Medium	High	Low	Medium	High				1st version	2nd version	1st version	2nd version
1.	6	10	0	23	9	2	29	19	2	1	9	60	71
2.	10	0	1	15	3	0	25	3	1	10	17	88	90
3.	6	1	0	3	4	0	9	5	0	7	10	90	92
4.	10	9	10	7	8	11	17	17	21	5	24	78	90
5.	0	2	3	11	2	1	11	4	4	2	7	70	85
6.	5	4	2	9	0	0	14	4	2	2	4	70	75
7.	3	3	1	0	2	7	3	5	8	7	10	85	87

[3] The focus of the present study is to examine DOP. Whereas Reactivity is controlled for, it is not being addressed as a variable. For a thorough review of the issue of Reactivity in SLA studies, the authors recommend reading Bowles (2010) and Yang and Zhang (2023).

8.	4	5	3	7	13	5	11	18	8	7	16	86	87
9.	4	6	2	5	11	6	9	17	8	8	13	76	79
10.	2	4	1	7	8	4	9	12	5	5	10	77	79
11.	6	3	0	8	10	7	14	13	7	4	12	88	89
12.	4	7	4	5	12	9	9	19	13	10	12	89	91
13.	5	8	3	4	7	3	9	15	6	14	23	91	92
14.	2	9	0	9	8	0	11	17	0	10	16	77	78
15.	6	11	2	10	5	6	16	16	8	7	17	79	90
Grand Total	73	82	32	123	102	61							

The results in Table 1 indicate that participants processed lexical items more at a medium level (82) than at a low level (73) or at a high level (32). This suggests learners spent more time processing lexical items according to the collected think-aloud data, which suggests that participants employed a greater level of cognitive effort to focus on the meaning of the target item. Most participants processed the grammatical items at a low level (123) rather than intermediate (102) or a high level (61). This could suggest that participants' level of cognitive effort was lower for grammatical items than for lexical items. However, if we just attend to the high DOP, there were more instances produced for grammatical items than for lexical ones.

4.2.Research Question 2

In an attempt to find out how DOP was related to changes in the written product for participants thinking aloud (n=15), the accuracy and quality of the first and revised versions were compared. For this, a paired sample t-test was run in order to discover how DOP is related to the changes in the essays of participants in the group thinking aloud (n=15). There was a significant difference when comparing both accuracy: $t(14)= -6.10$, $p= .00$, and quality: $t(14)= -3.75$, $p= .00$, in participants' performance in the pre and post-test. Descriptive statistics can be seen in Table 2.

Table 2. *Descriptive Statistics for Accuracy and Quality for the first and second versions*

		Mean	Std. Deviation	Std. Error Mean
Pair 1	Accuracy Version1	6.53	3.68	.950
	Accuracy Version2	13.33	5.55	1.43
Pair 2	Quality Version1	80.26	8.97	2.31
	Quality Version2	85.00	6.81	1.75

The paired samples t-test analysis reflects only the performance of the group thinking aloud, but not that of the silent group. Two additional analyses, independent t-tests, were conducted for both Accuracy and Quality in the second version to ensure that thinking aloud was not reactive and did not impact, positively or negatively, participants' performance in the condition thinking aloud when comparing it to those in the silent group. Results showed no

reactivity for accuracy, t(22.48)= .16, p= .06, or for quality, t(26.03)= 2.21, p= .14, indicating that thinking aloud did not significantly alter the TA group performance. Descriptive statistics can be seen in Table 3.

Table 3. *Descriptive statistics for TA and Silent conditions of Accuracy and Quality in versions 1 and 2.*

	Groups	Mean	Std. Deviation	Std. Error Mean
Writing Accuracy V1	Teacher FB+TA	6.53	3.68	.95
	Teacher FB+silent	5.40	2.26	.58
Writing Accuracy V2	Teacher FB+TA	13.33	5.55	1.43
	Teacher FB+silent	13.60	3.22	.83
Writing Quality V1	Teacher FB+TA	80.26	8.97	2.31
	Teacher FB+silent	73.93	5.54	1.43
Writing Quality V2	Teacher FB+TA	85.00	6.81	1.75
	Teacher FB+silent	80.13	5.13	1.32

5. DISCUSSION

The present study sought to examine the role that Depth of processing plays with regards to the application of corrective feedback on EFL learners' essays. Using Leow's operationalization model of DOP (2015), each participant's level of processing for lexical and grammatical items was measured. Table 1 presented each participant's level of processing (low, intermediate and high) regarding lexical or grammatical items before and after receiving teacher feedback.

The results show that participants processed the lexical items mostly at a medium level, rather than at other levels of processing, whereas the grammatical items were processed mostly at a low level, rather than at medium or high levels. The medium level of processing suggests that a greater amount of time was spent when processing lexical items than grammatical ones and participants employed a greater level of cognitive effort to focus on the semantic features of the target item (Craik & Lockart, 1972; Craik & Tulving, 1975; Leow & Mercer, 2015). An example of participant (number or letter) processing a lexical item while thinking aloud is presented below:

(1) *...Parks and green spaces can affect people's spirits very much regarding* ***to*** *urban life, industrialization of society and technology development....*

اینجا *to* نداریم. میخواستم بگم با توجه به زندگی شهری ... *people's spirits* ... پارک و فضای سبز می تونه تاثیر زیادی داشته باشه... می تونه تاثیر بذاره رو ... روحیه ی مردم با توجه به زندگی شهری ... بعد صنعتی شدنش ... اینو مثلا مثل *pay attention to* اینم میخواستم بگم *regarding to* ولی واسه *regarding* پس *to* استفاده نمی کنیم ... *parks and green spaces can affect people's spirits very much ...regarding urban life*

(We don't use to here. I wanted to say regarding urban life... people's spirits... parks and green spaces can have a lot of effect ... can effect ... people's spirits regarding urban life ... then its urbanization ... I wanted to use regarding same way I use pay attention to, but therefore we don't use to for regarding... parks and green spaces can affect people's spirits very much regarding urban life...)

This example depicts a lexical mistake that was marked as (E= extra word) for the participant to attend to. According to Leow's operationalization model (2015), this participant made an accurate connection between meaning and form and showed a deep level of cognitive effort as she spent time thinking and analyzing her reason for using the extra preposition. She drew upon

her background knowledge and said she was applying the same preposition of *pay attention* for *regarding*, as she had considered them to have the same preposition. She provided a correct translation of the target item and found out a different way to convey the same meaning.

Although lexical items were mostly processed at high or intermediate levels of processing, few learners processed the grammatical items at a high level, suggesting that they did not indicate high potential for processing the target items according to grammatical rules, as they mostly read the target items out loud, expressed that they were not sure about the grammatical rules or the reason why they had made a mistake; either arriving at the correct or incorrect alternative. A shallow depth of processing of grammatical items suggests that participants did not spend much time analyzing them, thus, implying that learners merely paid attention to different grammatical features of words that had been marked in the corrections, employing a lower level of cognitive effort. Another participant, participant XX processed a grammatical item at a low level when he wrote:

(2) *...The number of cars go up day by day and it causes a lot of traffic jam and air* ***polluted****...*
خب اونجا من نوشته بودم *polluted* در صورتی که فکر میکنم باید *pollution* استفاده کنم.
(ok, I wrote polluted there but I think I should use pollution instead.)

The above transcription includes the thoughts of Participant XX when processing a grammatical mistake. The word *polluted* was marked as (WF=wrong format) for him. He showed no potential for processing this grammatical item as he read it quickly, did not translate it and left the target item in English. As he was quite sure of the alternative he provided, he did not spend much time processing it, thus he showed a low level of cognitive effort to process the target item, whereas an intermediate or high level of processing would require learners to spend more time analyzing and drawing upon their prior knowledge to figure out the reasons for their mistakes or to recognize the appropriate grammatical rules for correcting them (Leow & Mercer, 2015). Leow (2015, 2019) pointed out that shallow or low processing does not lead to memory traces whereas a deep or high level of processing is connected to a conceptual or semantic analysis, degree of attention, and elaborative analysis.

The findings of the present study are in line with VanPatten (1990), Craik and Lockart (1972), Craik and Tulving (1975), Morgan-Short, Heil, et al. (2012), Leow (2008) and Leow (2015). Learners processed the grammatical items (form) at a lower level whereas they focused on and analyzed the lexical items (meaning) at a deeper level (medium). The rationale behind lexical and grammatical items being processed at different levels might be due to VanPatten's argument adopted from cognitive psychology that learners have a limited attentional capacity while processing information (McLaughlin, 1987; McLaughlin, Rossman, & McLeod, 1983). He argued that L2 learners, particularly those at lower levels of proficiency, would have difficulty processing both form and meaning concurrently (VanPatten, 1990). Learners usually do not process lexical items deeply when the purpose of reading the text is basically focusing on form rather than understanding the meaning. However, when the aim is to process both lexical and grammatical items, learners tend to spend less time analyzing the form and focus more on the meaning or lexical items. A plausible explanation for this is that not fully grasping the grammatical items in a text will not prevent learners from processing the gist of a text, as long as they understand its overall meaning.

The results of the present study are consistent with Leow's (2008) in that learners paid attention to form while processing for meaning. Nevertheless, the same was not true for the simultaneous processing of form and meaning. Participants spent more time processing their lexical mistakes as they were trying to analyze the pointed-out problems in order to find the best way to correct them. Therefore, according to Leow's profile, they applied a greater cognitive effort analyzing

for meaning rather than a shallow effort, which is limited to the appearance of words (Craik & Lockhart, 1975; Leow & Mercer 2015).

In an attempt to find out whether DOP relates to the improvements in participants' essays, a significant difference between the accuracy level of the students' essays in the pretest and posttest stages was found. The difference between pre and post-test findings suggests that the feedback received from the teacher and the reflections extracted from the collected think-aloud data have been beneficial for the final product. These findings are in line with what Leow (2015) asserts with regard to DOP being related to attention, noticing, awareness, cognitive processing, levels of analysis, and prior background knowledge, specifically regarding robust retention of grammatical and lexical items after the writing experience. However, in light of the results obtained in RQ 2, it also needs to be remembered that the group thinking aloud did not outperform participants in the silent condition in accuracy or quality on the composition second version. Therefore, whereas the use of verbal protocols seems a very valid tool to gain insight into participants' minds to better decipher how they are processing information in the L2, we need to be cautious when claiming that using TAs will boost learners' performance. Finally, even though Reactivity was not a variable addressed in the present study, it could be argued that, given the current findings, the use of TAs in this study was not reactive. This could offer some practical implications for its employment in further studies that want to better understand learners' cognitive processes while completing an L2 task.

6. LIMITATIONS

As with any empirical research, this study is not exempt from limitations. The first one is related to the sample size. Whereas, ideally, a larger sample would have taken part in the study to improve the generalizability of the findings, the present size, 30, was limited to the students enrolled in the EFL course. Future studies could look to extend the present study design by having a larger participant size to see if the present findings still hold. The second limitation lies in the design itself. The present study included a pre post-test design, but not a delayed post-test to account for retention in the long term. Ideally, further studies interested in replicating or extending the present research should include a delayed post-test to better account for long-term retention. Finally, a third limitation derives from the use of Think-aloud protocols to gain an insight into participants' stream of thought. Whereas researchers trained participants on how to think aloud, the fact that they did not say aloud everything that came to their minds cannot be completely discarded.

7. CONCLUSIONS AND PEDAGOGICAL IMPLICATIONS

The present study was conducted in an attempt to explore the strand of research initiated by Leow et al. (2008) and further explored by Adrada-Rafael (2017), Kim and Bowles (2019) or Manchón et al. (2020), among others, addressing the benefits of focusing on DOP to better understand the cognitive processes of participants as they work to improve their writing ability in the L2. It aimed to examine the role of Depth of processing upon incorporating corrective feedback on the writings of EFL language learners and finding out how DOP is related to changes in the written product. The present findings suggest that students were inclined to process and analyze their lexical mistakes at a greater depth than their grammatical errors. Current findings also suggest that the accuracy and the quality of students' writings improved as a result of the provision of feedback as well as simultaneously thinking aloud about the lexical and grammatical difficulties in their essays. Analyzing their TA data revealed how learners were trying to discover better, more appropriate alternatives to improve the overall quality of their writing. The present findings imply that L2 practitioners could focus on DOP as a potential factor in improving students' essays in terms of lexical and grammatical items as well as their quality and accuracy. Language instructors can promote DOP by assigning "educational psycholinguistics-based tasks" that can guarantee learners are "cognitively engaged" while attending to these tasks. (Leow & Adrada-Rafael, 2018, p. 194; Leow &

Mercer, 2015, p.80). While making learners think out loud in the L2 classroom to reflect on their writing might seem a bit unnatural for L2 practitioners, it could be an activity carried out outside of the classroom, where L2 learners could feel more at ease completing it. The subsequent reflections derived from their thinking aloud could then be shared in the classroom, and follow-up tasks could be planned by the L2 instructor to ensure learners benefit from having processed the feedback related to their writing errors. In light of the present results, it might be tempting to firmly conclude that using indirect coded feedback and promoting students' DOP via verbal protocols has led to learning and improvement in the learners' essays. However, the findings laid out in this manuscript should be interpreted with caution, as only two essays were collected from participants. Future studies could add a third essay to observe whether the DOP effects are carried on in time to a third manuscript. Furthermore, learners' intermediate proficiency level in the L2 needs to be considered when interpreting the current findings. Adding participants with a higher or lower L2 proficiency could potentially offer different results and comparing more than one group with different proficiency levels could account for how this factor might affect DOP and subsequent revisions of a writing piece.

The findings reported in this empirical study have aimed to contribute to the existing body of literature in the area of DOP within the field of Instructed Second Language Acquisition (ISLA), and, more specifically, within the discipline of L2 writing. Further investigations in DOP and L2 learning will allow both L2 researchers and teaching practitioners alike to better understand, from a psycholinguistic perspective, how learners process the information provided to them, and how this processing impacts learners' performance at a subsequent L2 writing task. This understanding could then be translated into the creation of pedagogical tasks to be implemented inside or outside the L2 classroom with the ultimate goal of improving learners' writing skills in the L2.

REFERENCES

Adrada-Rafael, S. (2017). Processing the Spanish imperfect subjunctive: Depth of processing under different instructional conditions. *Applied Psycholinguistics*, *38*(2), 477–508. doi:10.1017/S0142716416000308.

Adrada-Rafael & Filgueras-Gómez (2019). Reactivity, language of think-aloud protocol, and depth of processing in the processing of reformulated feedback. In *The Routledge Handbook of Second Language Research in classroom learning* (Ed. R. P. Leow) (pp. 199-211). Routledge.

Bowles, M. A. (2010). The think aloud controversy in second language research. New York: Routledge.

Caras, A. (2019). Written corrective feedback in compositions and the role of depth of processing. In *The Routledge handbook of second language research in classroom learning* (pp. 186-198). Routledge.

Cerezo, L., Manchon, R. M., & Nicolás-Conesa, F. (2019). What Do Learners Notice While Processing Written Corrective Feedback?: A Look at Depth of Processing Via Written Languaging 1. In *The Routledge handbook of second language research in classroom learning* (pp. 171-185). Routledge.

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

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268.

Ekuni, R., Vaz, L. J., & Bueno, O. F. A., (2011). Levels of processing: the evolution of a framework. *Psychology & Neuroscience.* 4(3), 333-339.

Ferris, D. R. (2002).Treatment of error in second language student writing. Ann Arbor: The University of Michigan Press.

Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). Testing ESL composition; A practical approach. Rowley, MA: Newbury House.

James, C. (1998). Errors in language learning and use. Essex, UK: Addison Wesley Longman.

Kahraman, A. (2013). Affective and cognitive effects of coded teacher feedback on FL writing students. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi [Hacettepe University Journal of Education], 28*(1), 189-201.

Kim, H. R., Bowles, M. (2019). How deeply do second language learners process written corrective feedback? Insights gained from think alouds. *TESOL Quarterly. 0*(0), 1-26.

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

Leow, R. P. (2020). L2 writing-to-learn. *Writing and language learning: Advancing research agendas, 56*, 95-118.

Leow, R. P. (2019). The Routledge Handbook of Second Language Research in Classroom Learning. NewYork.Taylor & Francis.

Leow, R. P. (2015). Explicit learning in the L2 classroom. New York. Routledge.

Leow, R. P. & Adrada-Rafael, S. (2018). La Atención y la concienciación en el campo de la adquisición de segundas lenguas. In (Eds. I. Finger & E. Ortiz), *A Dinamica do Processamento Bilingüe*. Brazil: Pontes.

Leow, R. P., Hsieh, H. C., & Moreno, N. (2008). Attention to Form and Meaning Revisited. *Language Learning*, *58*(3), 665–695.

Leow, R. P. & Mercer, J. D. (2015) Depth of processing in L2 learning: theory, research, and pedagogy, *Journal of Spanish Language Teaching*, 2(1), 69-82.

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.

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

McBride, S., & Manchón, R. M. (2023). Analysing L2 writers' processing of written corrective feedback via written languaging and think-aloud protocols. *Research methods in the study of L2 writing processes*, *5*, 337.

McLaughlin, B. (1987). Theories of Second-Language Learning. London: Edward Arnold.

McLaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: An information processing perspective . *Language Learning*, 33, 135–157.

Morgan-Short, K., Heil, J., Botero-Moriarty, A., & Ebert, S. (2012). Allocation of attention to second language form and meaning: Issues of think alouds and depth of processing. *Studies in Second Language Acquisition, 34*, 659–685.

Poulisse, N. (1999). Slips of the tongue: Speech errors in first and second language production. Amsterdam: John Benjamins.

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.

Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition*, 29(1), 67-100.

Shizuka, T. (2000). Effects of different editing methods on EFL writing quality: What do the students think? *Kanto Koshinetu Eigo Kyoiku Gakkai*, *14*, 43-54.

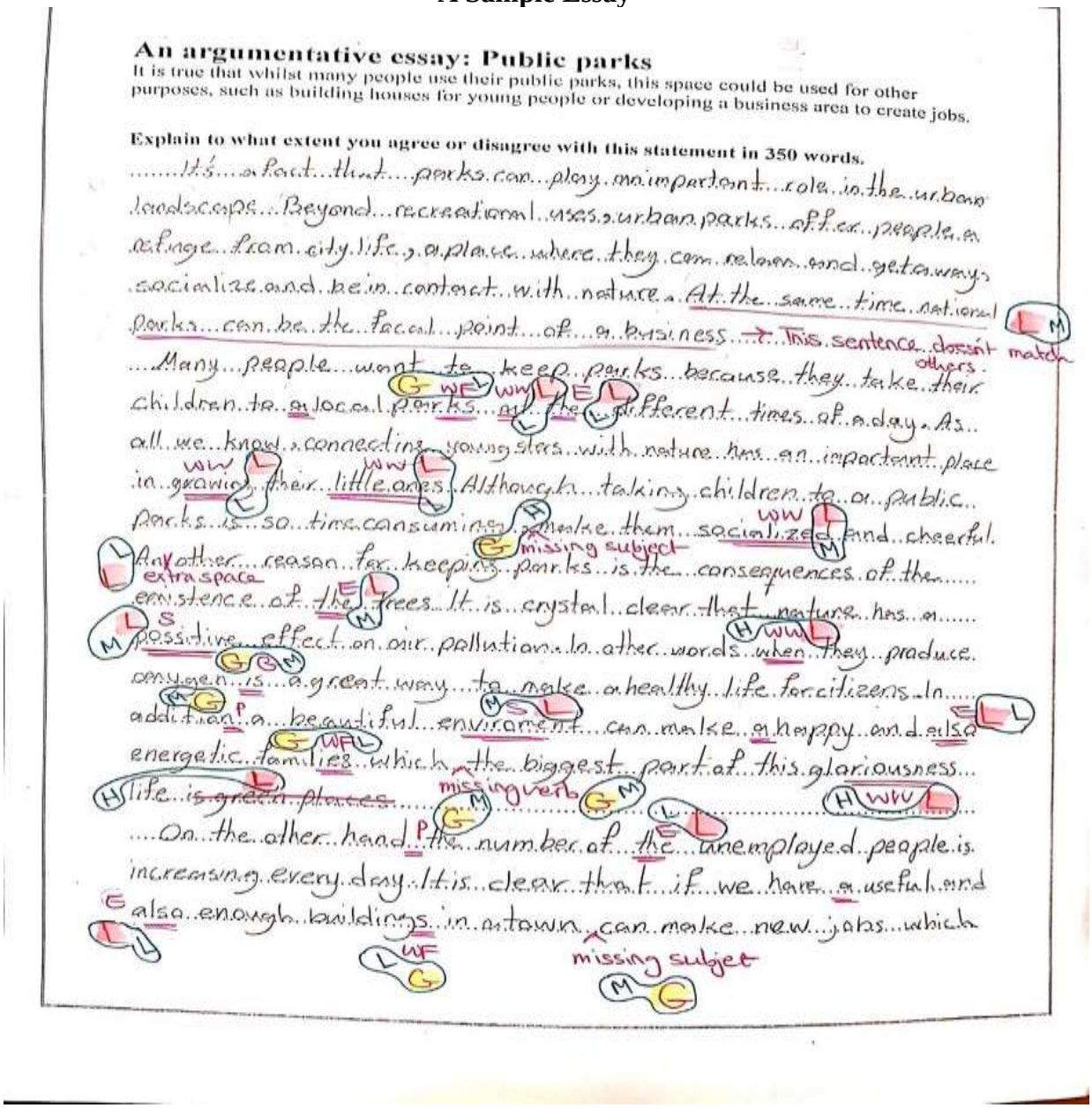
VanPatten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition*, *12*(03), 287–301.

Yang, C., & Zhang, L. J. (2023). Reactivity: Results from Quantitative Analyses of Performance. In *Think-Aloud Protocols in Second Language Writing: A Mixed-Methods Study of Their Reactivity and Veridicality* (pp. 87-138). Cham: Springer Nature Switzerland.

Yanguas, I., & Lado, B. (2012). Is thinking aloud reactive when writing in the heritage language? *Foreign Language Annals*, 45(3), 380–399.

Appendix A

A Sample Essay



An argumentative essay: Public parks

It is true that whilst many people use their public parks, this space could be used for other purposes, such as building houses for young people or developing a business area to create jobs.

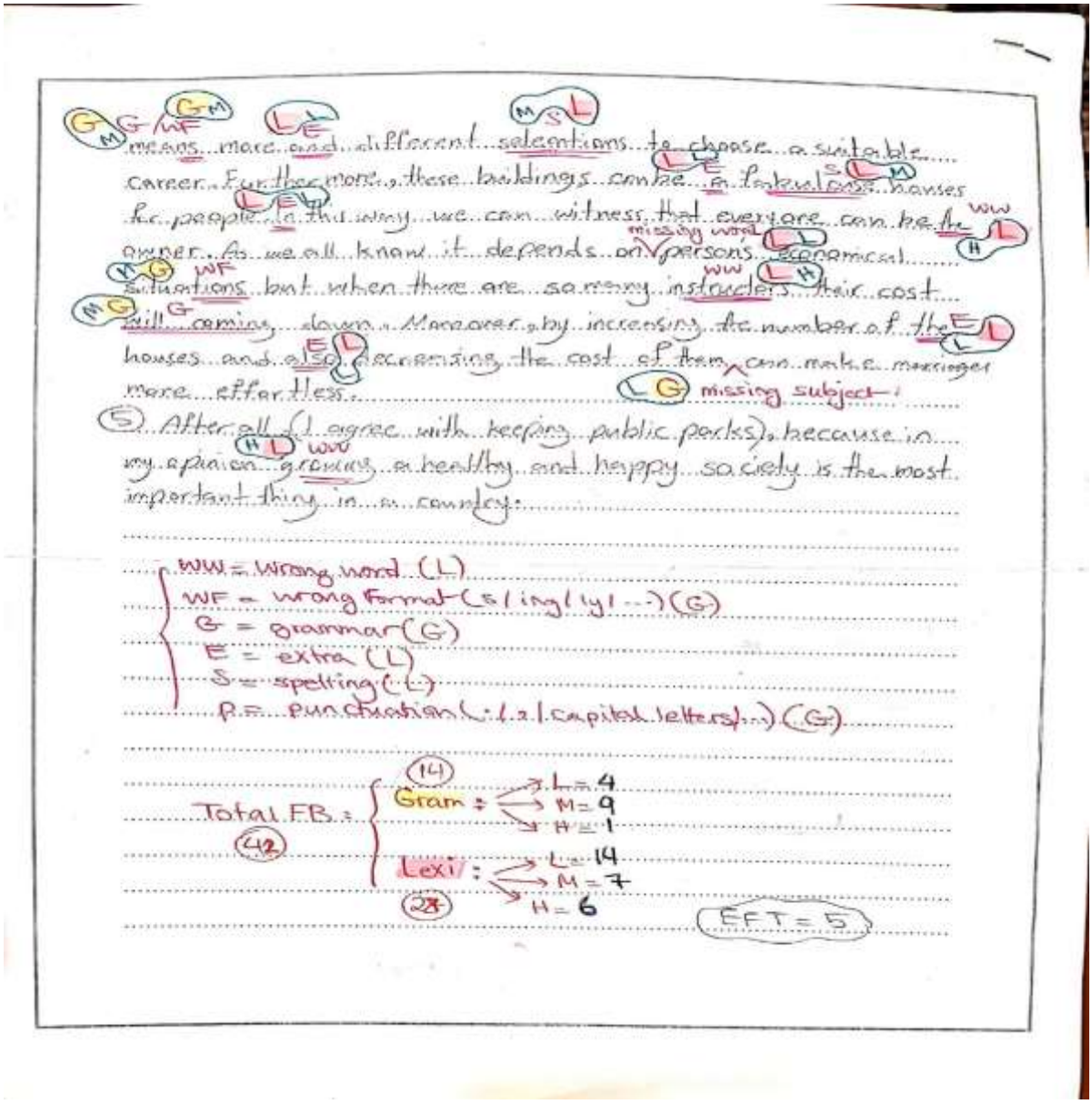
Explain to what extent you agree or disagree with this statement in 350 words.

It's a fact that parks can play an important role in the urban landscape. Beyond recreational uses, urban parks offer people a refuge from city life, a place where they can relax and get away, socialize and be in contact with nature. At the same time, national parks can be the focal point of a business. → This sentence doesn't match others.

Many people want to keep parks because they take their children to local parks at the different times of a day. As all we know, connecting young stars with nature has an important place in growing their little ones. Although taking children to a public parks is so time consuming, make them socialized and cheerful. missing subject

Another extra space reason for keeping parks is the consequences of the existence of the trees. It is crystal clear that nature has a positive effect on air pollution. In other words, when they produce oxygen is a great way to make a healthy life for citizens. In addition, a beautiful environment can make a happy and also energetic families which the biggest part of this gloriousness life is green places. missing verb

On the other hand, the number of the unemployed people is increasing every day. It is clear that if we have a useful and also enough buildings in a town can make new jobs which missing subject



Appendix B

ESL Composition Profile (Jacobs et al., 1981)

ESL COMPOSITION PROFILE			
STUDENT DATE		TOPIC	
SCORE	LEVEL	CRITERIA	COMMENTS
CONETENT	30- 27 26-22 21-17 16-13	**EXCELLENT TO VERY GOOD:** knowledgeable • substantive • thorough development of thesis • relevant to assigned topic **GOOD TO AVERAGE:** some knowledge of subject • adequate range • limited development of thesis • mostly relevant to topic, but lacks detail **FAIR TO POOR:** limited knowledge of subject • little substance • inadequate development of topic **VERY POOR:** does not show knowledge of subject • non-substantive • not pertinent • OR not enough to evaluate	

ORGANIZATION	20-18 17-14 13-10 9-7	**EXCELLENT TO VERY GOOD:** fluent expression • ideas clearly stated/supported • succinct • well-organized • logical sequencing • cohesive **GOOD TO AVERAGE:** somewhat choppy • loosely organized but main ideas stand out • limited support • logical but incomplete sequencing **FAIR TO POOR:** non-fluent • ideas confused or disconnected • lacks logical sequencing and development **VERY POOR:** does not communicate • no organization • OR not enough to evaluate	
VOCABULARY	20-18 17-14 13-10 9-7	**EXCELLENT TO VERY GOOD:** sophisticated range • effective word/idiom choice and usage • word form mastery • appropriate register **GOOD TO AVERAGE:** adequate range • occasional errors of word/idiom form, choice, usage but meaning not obscured **FAIR TO POOR:** limited range • frequent errors of word/idiom form, choice, usage • meaning confused or obscured **VERY POOR:** essentially translation • little knowledge of English vocabulary, idioms, word form • OR not enough to evaluate	
LANGUAGE USE	25-22 21-18 17-11 10-5	**EXCELLENT TO VERY GOOD:** effective complex construction • few errors of agreement, tense, number, word order/ function, articles, pronouns, prepositions **GOOD TO AVERAGE:** effective but simple constructions • minor problems in complex construction • several errors of agreement, tense, number, word order/function, articles, pronouns, prepositions but meaning seldom obscured **FAIR TO POOR:** major problems in simple/complex constructions • frequent errors of negation, agreement, tense, number, word order/function, articles, pronouns, prepositions and/or fragments, run-ons, deletions • meaning confused or obscured **VERY POOR:** virtually no mastery of sentence construction rules • dominated by errors • does not communicate • OR not enough to evaluate	
MECHANICS	5 4 3 2	**EXCELLENT TO VERY GOOD:** demonstrates mastery of conventions • few errors of spelling, punctuation, capitalization, paragraphing **GOOD TO AVERAGE:** occasional errors of spelling, punctuation, capitalization, paragraphing but meaning not obscured **FAIR TO POOR:** frequent errors of spelling, punctuation, capitalization, paragraphing • poor handwriting • meaning confused or obscured **VERY POOR:** no mastery of conventions • dominated by errors of spelling, punctuation, capitalization, paragraphing • handwriting illegible • OR not enough to evaluate	

Appendix C

Operationalization of Depth of Processing (DOP), Leow (2015)

Operationalization of Depth of Processing (DOP): Lexical Items

	LEVEL 1	LEVEL 2	LEVEL 3
	Low depth of processing	Medium depth of processing	High depth of processing
Description	Shows no potential for emerging form-meaning connection	Provides some evidence of processing target item	Provides evidence of making accurate form - meaning connection
Descriptors	reads target quickly translates the phrase to English but leaves the target in Spanish says s/he isn't sure what it is says s/he will click something repeats the target item carefully pronounces target word does not spend much time processing target item low level of cognitive effort to get meaning of target item	spends a bit more time processing target item makes a comment that indicates some processing of target item some level of cognitive effort to get meaning of the target item	spends time processing target item provides an accurate translation of the target item or finds a different way to say almost the same thing high level of cognitive effort to get meaning of the target item

Operationalization of Depth of Processing (DOP): Grammatical Items

	Low depth of processing	Medium depth of processing	High depth of processing
Level of awareness	*Noticing*	*Reporting*	*± Understanding (based on accuracy of underlying rule or form-meaning connection)*

Description	Shows no potential for processing target form grammatically	Comments on target item in relation to grammatical features	Arrives at an inaccurate, partially or fully accurate target underlying grammatical rule
Descriptors	reads target quickly translates the phrase to English but leaves the target in Spanish carefully pronounces target item repeats target item Says s/he isn't sure what it is does not spend much time processing the target	spends a bit more time processing the target item makes comments that indicate some processing of the target item some level of cognitive effort to process the target item grammatically	makes hypotheses regarding the target item provides an inaccurate, accurate and/or partially accurate rule corrects previous translation spends much time processing target items a high level of cognitive effort to process target items grammatically

About the Author

Sergio Adrada-Rafael, *Ph.D., is an Associate Professor of Spanish and Applied Linguistics in the Department of Modern Languages and Literatures at Fairfield University, USA. He teaches courses in Spanish language (Elementary and Intermediate levels), Hispanic Linguistics, Bilingualism, Spanish for specific purposes, and Psycholinguistics. His research has been published in peer-reviewed journals like Applied Psycholinguistics, the International Journal of Applied Linguistics, Languages, and the Euro-American Journal of Applied Linguistics and Languages.*

Rose Meschi *holds a Ph.D. in Applied Linguistics from the University of Tehran, Iran. Her primary research interests span generative AI, active learning, critical thinking, written corrective feedback, depth of processing, and academic writing. In addition to her research endeavours, she authored the book "English for the Workplace," published by Sultan Qaboos University Publications. She is currently serving as a Senior lecturer and an Assessment Unit member at Sultan Qaboos University.*