Task Manipulation of Planning Conditions: An Analysis of Syrian Online L2 Performance in Text-based & Voice chat

by

SH. ADI

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

The University of Sheffield

2022

Task Manipulation of Planning Conditions: An Analysis of Syrian Online L2 Performance in Text-based & Voice chat

by

Shahla Adnan Adi

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

School of English

February 2022

<u>Abstract</u>

This research was based on a mixed methods investigation of a longitudinal study that was conducted in the text-based and voice chat environments to report on the impact of task manipulation on a guided pre-planning/no-planning condition in connection with second/foreign language (L2) performance. Data were collected from two groups of ten intermediate level Syrian learners (SLs) as they were interacting online over a six-week period of text-based chat with either a native speaker of English (NS) with teaching experience or a Syrian English teacher (ST). Working in pairs, dyads performed a series of tasks that varied in type, complexity, and cognitive demands. The learners were placed into one of the following experimental groups: the guided pre-planning group or the no-planning group. Two months later, the same groups of learners experienced similar conditions and fulfilled similar language tasks, except for the fact that the interaction was carried out orally this time in the form of voice chat. Independent samples and paired samples analysis was carried out to measure any possible significant difference in terms of the complexity, accuracy, and fluency of learners' production in the immediate and delayed post-tests. The quantitative/qualitative analysis of chat exchanges, pre-/post-tests, and stimulated recall interviews provided some evidence as to the effectiveness of manipulating task type/complexity in promoting L2 proficiency. The findings revealed, however, that learners' progress was also moderated by other external factors, such as individual differences, motivation and anxiety levels. All the participants had positive attitudes towards the online experiences and perceived the online interaction as beneficial for L2 development, and the pedagogical implications of the findings for L2 English classrooms are discussed.

Table of Contents

List of Tables:

List of Figures:

List of Graphs:

<u>ACKNOWLEDGMENTS</u>

<u>DEDICATION</u>

This thesis is wholeheartedly dedicated to the memory of my beloved mother; Malak Al-Baroudi, who passed away January 2020. I had promised to make you proud and now I hope that I have fulfilled my promise.

DECLARATION

The opinions expressed in this thesis are solely those of the author and acceptance of this thesis as a contribution to the award of a degree cannot be regarded as constituting approval of all its contents by the school of Arts and Humanities. I certify that all the materials in this dissertation which are not my own have been identified and properly attributed.

Name: Shahla Adi

Signature: _____

Date: 28/02/2022

1. <u>Introduction</u>

In line with ongoing advances in technology, language learners have found the space to extend their learning opportunities beyond the limits of place and time inherent in traditional classroom face-to-face (F2F) interaction. In this vein, Otto (2017) claims that L2 learning technologies and computer-assisted language learning "evolved from delivery via localized technological resources to any-time, any-place provision through networked digital tools" (cited in Tavakoli et al., 2019: 6). Considerable attention has been devoted to the role synchronous computer-mediated communication (SCMC) plays in promoting second language acquisition (SLA) (e.g., Blake & Zyzik, 2003; Mackey, 2012; O'Rourke, 2005; Sauro, 2009; Smith, 2003). SCMC is a means of human interaction that comprises a variety of comprehension and production activities. What seems to have led to a particular surge of interest in this topic has been the fact that this medium could afford perfect "learning conditions to support both meaning-oriented communication and focus-on-form reflection", which in turn are crucial for L2 development (Lee, 2008: 53). The burgeoning interest in enhancing L2 proficiency via text based SCMC has led to a proliferation of various studies that with very few exceptions have systematically examined the efficacy of chatroom environments in developing oral competencies in particular (e.g., Blake et al., 2008; Chun, 1994; Ko, 2012; Lys, 2013; Payne & Whitney, 2002; Sauro, 2012). Additionally, Gonzalez-Lloret (2017) contends that task-based language teaching is the most efficient approach to fully apprehend the potentials of technological tools for L2 learning and L2 development. Relatively little research, however, has investigated the impact of task type (Aydin & Yildiz, 2014) and different planning conditions (Hsu, 2017) on L2 development. Even fewer studies have inspected the effects of online interaction on learners' anxiety levels (Satar & Ozdener, 2008). At the time of writing, none of the previous studies have examined the combined effects of manipulating task type/complexity in connection with a strategic pre-planning condition over time in two

1

different modes of SCMC (i.e., text-based and audio chat). Additionally, the Syrian context, where English is taught as a foreign language and learners rarely have any chance to practise their English inside, let alone outside the classroom, has never been investigated in terms of learners' proficiency in the SCMC modality. In my experience as a Syrian English language teacher, this lack of practice inside Syrian language classrooms could be tied to (i) large class sizes; and (ii) a predominantly grammar-translation approach, which means that very few opportunities for spoken interaction arise. Therefore, the current study usefully adds to previous SCMC research and provides particularly interesting findings as any improvements in learners' performance will be predominantly due to their online interaction; that is to say, the improvements will not be due to them practising English in their out of classroom environment. It is also worth mentioning that the findings of previous SCMC studies were mainly based on quantitative rather than qualitative analysis, leading to an incomplete picture as to the effectiveness of the online mode in promoting L2 proficiency. Thus, the overarching objective of this research is to build upon previous research in SCMC and provide a longitudinal analysis (i.e., more than fifteen weeks' treatment) on the effects of manipulating some features of interaction within a task-based learning approach on learning outcomes, attitudes, and L2 development, adopting a mixed-methods approach. Basically, I will focus on what ways the three dimensions of production, i.e., complexity, accuracy, and fluency (or CAF), compete for attention during task performance in synchronous text-based and audio chat interaction.

To this end, two separate studies, each incorporating a pre-test/post-test design, were carried out over the duration of five consecutive months. Different cognitively demanding tasks (e.g., narrative and decision-making tasks) were implemented during twelve weeks of text-based and voice chat sessions. The data generated from these two studies will provide valuable input for curriculum designers in their quest to determine task types and means of communication that

are likely to better foster L2 performance. A combination of morphological and syntactic measures was used to gauge the quality/quantity of the language produced by SLs in both studies (e.g., the average number of AS-units, error-free clauses, number of pauses). In the first study, SLs were asked to undertake a pre-test to assess their proficiency level in the aforementioned measures. Twenty SLs of an intermediate level were assigned accordingly to work with either NSs or STs under either the planning or no-planning group for six weeks (one session per week for each dyad) and perform different task types via text-based chat. The first group completed the tasks under what is called the guided pre-planning condition; this consists of step-by-step guidance towards using particular forms entailed in the task design while trying to maintain meaning in the task as a whole. The second group, however, acted as a control group in the current study in the sense that learners were given no time to plan their online performance and instead asked to start working on the task immediately. Having completed all the online sessions, SLs then completed an immediate post-test and a delayed post-test (a month later) to detect any short-term/long-term gains (i.e., uptake) in their L2 production. A similar routine was followed one month later with the same groups of learners in the form of audio chat interaction.

This research design is promising as it will reveal whether SLs' performance or their planning strategies are influenced by task type/complexity, planning time, and mode of interaction. It further pinpoints any long-term changes (e.g., gains) in learners' L2 production. The rationale for enrolling NSs and STs in the current study by no means aims to investigate whether the complexity of SLs' production would differ in line with the L1 or L2 status of their interlocutor. It is rather an attempt to offer SLs varied opportunities to benefit from the online interaction. Additionally, the time interval between the two studies (one month) aims to lessen the effects learners' performance in the first study might have on that of the proceeding one. Stimulated recall interviews were conducted with SLs immediately after each chatting session. Hence, as

the learners were reviewing the chat log, they were prompted to talk about interesting segments of their online interaction (e.g., instances of deletion, pausing, self-correction) and planning strategies. In addition, SLs were required to write a weekly report to evaluate and reflect on their online performance. This uncovered important information regarding learners' awareness of the progress they were achieving in their L2 production. Perceptions of all the participants towards the online experience as a whole were also elicited so that they could provide useful insights for researchers, teachers, and those responsible for curriculum development.

2. <u>Literature Review</u>

The purpose of this chapter is to discuss previous research on L2 development in SLA and SCMC contexts. It begins by looking at the features and affordances of online synchronous interaction (text-based and audio chat). Then it moves to consider the components of L2 proficiency, oral proficiency in CMC studies, written proficiency in CMC, and learner factors (e.g., anxiety levels). Perceptions and attitudes towards SCMC will be addressed afterwards. The final section will set out the aims as well as the research questions of the current study.

2.1 SCMC: features and affordances

Research examining the unique benefits of SCMC, defined as "real-time, synchronous conversation that takes place online", has proliferated and began in the 1990s (Baralt & Leow, 2016: 200). SLA scholars and practitioners have recently demonstrated a keen interest in researching how the affordances of the SCMC modality can be better utilized to facilitate L2 acquisition (Mackey, 2012). Now, despite the fact that SCMC differs from F2F communication, there were studies that drew some parallels between the two, including the aspects of real-time interaction and short turns, as well as the informality of discourse moves (e.g., Yilmaz & Yuksel, 2011). Therefore, Pellettieri (2000) contends that "language practice through [CMC] will reap some of the same benefits for second language development as practice through oral interaction" (p. 59). As evidence of this, Sauro (2012) examined learners' performance to compare the syntactic and lexical complexity of learners' output in the F2F and SCMC modes during the completion of narrative monologic tasks. The results indicated no significant difference in either the lexical or syntactic complexity generated in both mediums. In the same vein, Lin (2015) argues that previous SLA research seems "to endorse the use of CMC in the language classroom given that it creates a social context that is similar to face-to-face in which most features of authentic interaction can be replicated" (p. 263). SCMC has thus been seen to

offer learners heightened opportunities to work collaboratively, tackle communication problems instantaneously in real-life communication, and subsequently foster the cognitive processes necessary for L2 development.

However, most research on this area has usually been concomitant with identifying several advantages of online interactions over traditional classroom exchanges. For example, some of the earlier studies to examine the impact of online interaction compared to that of the F2F environment on L2 performance revealed more complex learner production in SCMC (Kern, 1995; Warschauer, 1996). Another reported advantage of the SCMC modality compared with F2F communication is that, in FL classrooms, where a large number of students are attending the same class, it is implausible that they will all get plentiful opportunities to practise their English. This could be due to time constraints and the high anxiety levels F2F spoken interaction may cause, especially in the case of reticent students. Yet, heralded for generating a less stressful context for L2 practice, CMC environments could result in an increased willingness on the part of shy learners to take risks and try out new hypotheses in their output (Adams et al., 2015). Besides, online synchronous chatting tools could be viewed as an ideal substitute when F2F discourse is not practical or does not function as expected. This could often be the situation with large classes or contexts where interaction in L2 is not possible outside the classroom (Blake et al., 2008; Satar & Ozdener, 2008).

The preceding discussion indicates that CMC as a venue for human interaction can prove useful as a complement to F2F dialogue. The following sections further highlight the potential benefits of two modalities of SCMC (i.e., text-based & voice/audio-chat) that will be deployed in the current research.

2.1.1 Text-Based Chat

Text chat refers to real-time typed messages, a new evolving hybrid form of electronic interaction that carries many of the same features of both spoken (i.e., short forms, contractions)

as well as written discourse (i.e., formal and complex output). Pellettieri (2000) contends that due to the synchronous nature of its messages, text chat bears a striking resemblance to oral interaction. That is, it might enhance learner opportunities for language development, such as providing affordances for negotiation of meaning (NofM)[1] and repair moves. Results of more recent studies, however, showed that instances of NofM in SCMC are qualitatively different from those of F2F. That is to say, in the former, learners produce linguistically more complex and more modified utterances (Moradi & Farvardin, 2020), especially when they initiate modifications and self-repairs (Smith, 2008). The importance of negotiating for meaning in the discourse of L2 students has been highlighted by Long (1996) as it serves to make learners aware of the disparities or the variances between their production and that of the target language. Varonis & Gass (1985) introduce a model that illustrates how the discourse structure unfolds during negotiations. They describe the sequence of the negotiation routines as follows: Trigger (T) →indicator (I) →response (R)→reaction to response (RR).

Table 1: Discourse Model of NofM[2]:

Utterance	Function
Sinister: now how do u think about the tuition?	Trigger
Maryam: tuition? I don't know its meaning	Indicator
Sinister: the money that we pay to study at school	Response
Maryam: yeah thanks	Reaction to Response

---

[1] NofM refers to "the modification and restructuring of interaction that occurs when learners and their interlocutors anticipate, perceive, or experience difficulties in message comprehensibility" (Pica, 1994: 495). Accordingly, NofM is triggered when a breakdown in the flow of a conversation occurs, leading to a modified utterance either from L2 learners or their interlocutors with the intent of ensuring shared understanding.

[2] Example adopted from Shekary and Tahririan (2006: 562).

As **Table 1** shows, the negotiation sequence begins when the speaker produces an utterance, known as a *trigger* (T), which in turn results in an *indication* (I) of non-understanding, i.e., questions or comments on the part of the hearer that something was not clear regarding the preceding utterance. The speaker then *responds* (R) by either repeating the message or providing a modified version of the utterance. The final prime, commonly referred to as '*reaction to the response*' (RR), is optional. In other words, it might or might not be part of the negotiation sequence, but if produced, it would tie up the routine and demonstrate that an uptake of some sort has already taken place.

However, even with some similarities and resemblance between traditional F2F interaction and that of text-based chat (e.g., real-time nature, repair moves), not all aspects of these conversational contexts are alike. In the text-based mode, for example, the slow rate of dialogue, or what Beauvois (1998) refers to as conversation in slow motion, affords learners some time, albeit relatively short, to react to their production as well as that of their interlocutors. In the same vein, Payne & Whitney (2002) argue that text-based CMC provides learners with an environment to practise language production at a reduced rate by "developing the same cognitive mechanisms underlying spontaneous conversational speech" and reducing memory load (p. 2). Hence, learners can engage in self-correction, have the chance to shape the form, and contemplate the content of their responses, all of which may promote significantly more complex and accurate utterances than their F2F equivalents (Cho, 2011; Sauro & Smith, 2010). Sauro & Smith (2010) investigated the syntactic complexity and lexical diversity of the language produced by the learners during text-based chat. Video recorded chat logs of 12 beginner-high level dyads of German learners, working on a sequential ordering task, were analysed accordingly to compare the overt (i.e., sent) and covert (i.e., deleted) segments of their production. Sauro & Smith distinguished three types of output: firstly, there was pristine output (overt) that learners typed and sent immediately with no deletions of any

kind; secondly, there was deleted output (covert), which appeared only in the chat recordings of the learners as they did not send it to the other interactant; and finally, there was post-deleted output (overt) that involved all the modified messages which followed the deleted texts. Results revealed that the post-deleted production, which demonstrated evidence of post-production monitoring, showed significantly more linguistic complexity and lexical diversity than the other two categories (i.e., sent and deleted output). Hence, regardless of their language proficiency level, one could argue that the slower speed of text-based online interaction compared to F2F helped learners attend to their production and accordingly produce more complex language. Presumably, an analysis of chat logs produced by higher-level L2 learners interacting with NSs or L2 teachers (as in the current study) would result in more robust effects than those found in Sauro and Smith. In other words, the learners might benefit from the feedback and the language produced by NSs and more advanced L2 speakers. Besides, conducting recall interviews rather than relying primarily on recorded chat logs could have also uncovered vital details as to the factors and conditions that influenced learners' decisions of whether to modify or send the message immediately to their partners.

The non-ephemeral nature of text chat—the possibility to view/review the messages on the screen by scrolling the cursor backward/forward—not only helps to free up learners' memory to notice and tackle more critical communication problems (Yilmaz & Yuksel, 2011), but also has a positive impact on enhancing their depth of processing, allowing quick hypothesis testing (i.e., comprehension and retention of received knowledge). This feature may motivate less competent L2 users to utilize new forms in their production afterwards. To put it simply then, the visual saliency of some complex structures as projected on the screen may facilitate the noticing of these forms (Gurzynski-Weiss & Baralt, 2014; Bower & Kawaguchi, 2011; Lai & Zhao, 2006), bringing some stimuli into focal attention (Mitchell et al., 2013), and consequently "help learners to either confirm or disconfirm currently held hypotheses about

the target language" (Sauro, 2009: 101). Lai & Zhao (2006) provide some preliminary support for this claim and illustrate how text chat enhances learners' noticing of problematic grammatical forms as well as the feedback they receive from more advanced interactants. Quantitative and qualitative analysis of their data revealed that text-based online sessions promote more noticing than the traditional F2F conversation, especially concerning learners' linguistic mistakes. This result was explainable given the longer processing time and relative permanency of interlocutors' utterances, both of which are inherent in text chat compared with the F2F mode. Yuksel & Inan (2014) also examined whether the communication mode (i.e., F2F vs text-based SCMC) had any effects on the instances of NofM generated during the interaction of 64 EFL learners, who completed two-way information-gap tasks (one in each medium), as well as the level of noticing of NofM by the learners based on a stimulated recall protocol. The findings revealed that although the average number of NofM routines was higher in the F2F exchanges, the SCMC group reported significantly more incidents of noticing of these routines than the F2F group. The participants who worked in the online mode also identified a higher average of communication breakdown stemming from lexical and grammatical items. This difference between the F2F and SCMC groups was found to be statistically significant.

The notion of saliency and noticing was originally proposed by Schmidt (1990, 2001), who postulates that SLA is conscious. Therefore, the provision of both positive and negative evidence is one of the ways to enhance learners' noticing of particular linguistic forms. According to Sotillo (2005), "positive evidence refers to the provision of grammatical utterances or well-formed statements in the learners' linguistic environment, whereas negative evidence implies the provision of either implicit or explicit corrective feedback" (p. 468). Hence, the latter serves to direct learners' attention to the differences between their output and that of the target language (Long, 1996). More recently Yilmaz (2016) also defined corrective

feedback as "the reactions language learners receive from their interlocutors indicating that the learners' language production is not target-like" and therefore, it is thought to assist L2 learning (p.1). Explicit feedback could be presented in the form of metalinguistic correction or any other type which overtly indicates that learners' output is not acceptable. Implicit feedback, however, involves recasts and modification devices such as clarification requests, comprehension checks, and other forms that do not overtly draw learners' attention to their erroneous production. There are indications that such negative evidence is crucial as it facilitates the acquisition of certain language forms, especially those that may be hard to learn through input alone. Thus, CF can be useful in acquiring complex structures by increasing the probability that they will be noticed. Likewise, negative evidence might also be beneficial in obtaining forms that are low in their perceptual salience, such as articles and third-person singular. These forms seem difficult for learners to perceive as they do not typically lead to communication breakdown when errors occur (Dekeyser, 2005). According to Gass (1997), enhancing the saliency of these forms may "help ensure that particular forms are noticed by the learner and hence lead to rule strengthening" (p.19). **Excerpts 1** and **2** below illustrate an example of explicit and implicit correction, respectively.

Excerpt 1: metalinguistic feedback (Sotillo, 2005: 479-481)

A: about the movie, I was a little confuse of who was the Matrix

L: You have to pay attention to your past tense, okay.

A: yes

Excerpt 2: (recast)

T: Yes, we're all from Guadalajara, but my father got a good job opportunity in Zacatecas which is four hours far away**.**

 D: Four hours away.

It stands to reason therefore to claim that CF delivered via online text-based chat, in particular (due to the time span, permanency of the text, and the saliency of its messages when projected on the screen compared to F2F or audio/video chat modes), holds potential advantages for L2 learning and for the acquisition of hard-to-learn or low-salient forms (Baralt, 2013; Chen & Eslami, 2013; Gurzynski-Weiss & Baralt, 2015; Henderson, 2021; Lin et al., 2013; Sauro, 2009; Sotillo, 2005; Zeng & Takatsuka, 2009; Yilmaz & Yuksel, 2011). Yilamz & Yuksel (2011), who investigated the role of the communication mode (F2F vs text-based SCMC) in the extent to which learners benefit from recasts, found that learners scored statistically significantly higher on receiving recasts through text-based SCMC than recasts through F2F communication. Sotillo (2005), on the other hand, previously examined CF provided in different combinations in SCMC. A total of six NS-NNS and NNS-NNS dyads worked on communicative and problem-solving activities over the duration of nine weeks. NSs and highly competent NNSs (who were enrolled in an undergraduate course required for ESL teacher certification) were not informed to focus exclusively on correcting learners' errors or to favour one or the other type of negative feedback (i.e., explicit and implicit). Instead, they were only encouraged to deliver comprehensible and meaningful input to their interlocutors who were not enrolled in ESL classes. A one-way chi-square analysis revealed that 65 of the total 159 opportunities for error correction and negotiation among interlocutors resulted in actual error correction. Lexical errors represented only 38% (25/65) of all error types detected in learners' production in both dyads' types. Missed opportunities to provide feedback on NNSs' erroneous utterances were traced back to participants' involvement with task completion. The number of actual corrections was higher in NNS-NNS dyads than in NS-NNS dyads (46 and 19, respectively). Of all correction episodes, 37% resulted in successful uptake (i.e., instances where learners incorporated the targeted language forms into their L2 production immediately following CF or afterwards during negotiation work). Sotillo also found that NSs provided

more indirect CF than highly competent NNSs. Since no interviews were conducted with the participants to seek answers as to why they favoured one type of feedback over the other, there was a lack of explanation of such results in Sotillo's study. However, this could be traced back to NSs' unfamiliarity with the learners, i.e., they possibly did not want the learners to feel intimidated by the explicit correction. Another plausible explanation that might affect NNSs' inclination to be more direct in their feedback (albeit they did not know their interlocutors) is the fact that they came from the same ESL context as the less competent NNSs. That is to say, they were used to learners' needs and learning styles. It is also possible that the enrolment of the highly competent L2 speakers (prior to the study) in a course required for ESL teacher certification might have affected their decisions about the type of feedback necessary for particular situations. During their training they became familiar with the distinction made in previous research between explicit and implicit feedback, especially when addressing non-salient or hard-to-learn grammatical forms.

Contending that meaning-focused instruction alone may not lead to the development of high-level accuracy, Chen & Eslami (2013) more recently examined the occurrence and impact of incidental focus on form in promoting L2 development in text-based chats. Sixteen college-level Taiwanese EFL learners (with an intermediate proficiency level) were partnered with sixteen native English speakers to complete two communicative tasks via Instant Messenger. Chen and Eslami assumed that working on meaning oriented tasks with more proficient language users who would be offering feedback to their interlocutors regarding their language use would help the learners alternate their attention between form and meaning and eventually facilitate L2 development. Following the orientation phase which lasted for three weeks and involved technological training and ice-breaking between the interlocutors, each dyad engaged in text-based chat for 90 minutes per week over the duration of six consecutive weeks to complete either a jigsaw or a decision-making task. These were followed by an immediate post-

test in week ten and a delayed post-test in week 14 to elicit the EFL learners' FoF (i.e., focus on form) memories and detect any short/long-term gains. Language-related episodes (LREs[3]) were identified and utilised as a basis for individualised tailor-made tests in order to assess learners' learning outcomes. The results of two post-tests showed that FoF in text-based chats was highly correlated with subsequent L2 development; a total of 425 LREs were analysed, and the findings showed that the participants were able to correctly recall and reproduce 70% of the target linguistic items in the immediate post-test and 69.39% in the delayed test. The number of LREs also implied that, unlike the awkwardness that often exists in F2F classrooms, the EFL learners felt comfortable to ask questions and express themselves during the online interaction.

Much-needed light should be shed, however, on how different modalities of SCMC, as well as different interactants (i.e., NS and competent NNSs), can affect various aspects of production and assist language development. The current study, therefore, will expand on this line of research. Later in this chapter, more literature will be reviewed as to the efficacy of text-based chat in enhancing L2 written proficiency. However, the following section will explore the interactive affordances of voice/audio chat.

2.1.2 Voice/Audio Chat

Voice chat refers to any oral synchronous conversation that takes place online, possibly using headphones or speakers. Yuangus (2010) refers to the need "to practice and foster communicative habits that can provide learners with the right tools to effectively communicate orally be it with other L2 learners or with native speakers in the outside world" (p.73). Indeed,

---

[3] LREs: instances where "learners show evidence of a shift of attention from message meaning to message form regardless of whether this shift emerges due to a problem with message comprehensibility" (Yilmaz, 2011:117).

the speed of interaction in the voice mode allows far more L2 production (and consequently more opportunities for learners to negotiate for meaning) in a rather shorter time when compared with written exchanges. Bueno-Alastuey (2011) postulates that oral SCMC (audio/video) shares many similarities with F2F conversation. It seems rational, therefore, to assume that negotiations produced via this mode bear a stronger resemblance to F2F discourse than those triggered during text chat. In other words, miscomprehension in the oral mode is likely owing to pronunciation, lexical, or speed issues as well as grammatical misunderstandings. Additionally, voice chat is thought to offer practice opportunities inside/outside the classroom that enhance learners' speaking/listening skills and consequently improve communicative competence and develop L2 oral proficiency (e.g., Bueno-Alastuey, 2013; Kost, 2004; Yanguas & Bergin, 2018). That is, the language user will practice the language in an authentic spontaneous unplanned context and will undergo the same cognitive pressure (as in F2F conversation) to process received input and formulate well-controlled, complex, fluent output, leading to non-trivial gains in oral proficiency. Eventually, it seems likely that L2 learners will not only exhibit reduced anxiety levels when speaking the language, but they will also showcase more willingness to ask questions, contribute to the conversation, and freely express their opinions. Bueno-Alastuey (2013), who studied interactional feedback in voice chat where learners focused on form and meaning simultaneously, concluded that this medium of SCMC is "a fertile ground for negotiated interaction and for making learners use their linguistic resources, so that they could modify their production and advance in their L2 learning process" (p.551).

However, unlike F2F communication, as learners do not have access to gestures or facial expressions during online interaction to better comprehend their partners, they primarily rely on language to deliver meaning and discuss instances that cause misunderstanding. To put it another way, learners need to produce more elaborated output while trying to employ different

communication strategies in rapid turn-taking interaction (Granena, 2016). Hence given all the aforementioned benefits, oral SCMC may promote interactions among language learners and get them ready for real-life F2F communication. Nevertheless, this tool of SCMC has been underresearched and only sporadically investigated despite the developmental role it might have on L2 oral proficiency. Thus, its applications within the language classroom and pedagogical implications for L2 learning remain untapped in many classrooms around the world (Yanguas & Bergin, 2018).

Studies that have examined online oral performance have adopted entirely different perspectives and research designs, making it difficult for future researchers to draw any general conclusions on the effectiveness of this tool for L2 development. Jepson (2005), for instance, compared differences in the types of repair moves (e.g., NofM and CF) generated during ten sessions of five minutes of text and voice chat using chat rooms on the Internet for five different days. Participants, six EFL learners using text chat and three voice chat, enrolled in an online school and used nicknames during the interaction; this anonymity was intended to make them more willing to engage in the online environment. Findings obtained from one-way chi-square tests revealed a significantly higher number of repair moves in the audio chats than in the text-based chats. Qualitative analyses of instances of NofM and CF also revealed that repair moves in the audio mode were mostly associated with a pronunciation issue. Yet, it was difficult to determine whether the repair moves were simply due to some technological problems (i.e., poor Internet connection). Besides, the nature of the chatroom interaction and the equipment used to record the online text chat did not allow the researcher to observe whether the participants edited their messages before sending them. That is, the researcher did not have access to the learners' screen and could only view these messages following their delivery to the other interlocutor. Therefore, it was not possible to measure particular repair moves, such as self-corrections, and thus potential data were missing from the analyses. Besides, as no

testing was done beforehand, the lack of information regarding learners' proficiency levels and language background made it hard to pinpoint or determine their initial differences and pair them accordingly.

Yanguas (2010), on the other hand, posits that previous studies have not fully addressed more important questions regarding whether research on voice chat promotes L2 acquisition, and how this mode could be best integrated into the language classroom so that it leads to positive effects on L2 development. Thus, in an attempt to provide further insight into some of these questions, he examined the interaction of intermediate level learners during task-based oral chat to analyse how they negotiate for meaning when working within the video and audio CMC groups. The study also aimed to scrutinize any possible differences between the online oral mode and F2F communication and whether the negotiation routines of the former are similar to those reported in the text chat literature. Fifteen dyads of Spanish learners were randomly assigned to a voice group, a video group, and a F2F control group. Dyads were asked to complete a jigsaw task that triggered the use of sixteen unknown lexical items. Data analysis of all chat transcripts showed instances of NofM during the online oral performance, mainly when non-understanding occurred between speakers. The lack of visual contact in the audio group resulted in different negotiation routines from the video groups, whereas no such difference was recognized between the video and F2F groups. Additionally, turn-taking patterns in the oral online chat were shown to be very similar to F2F conversation, unlike those found in the written mode[4] of SCMC. It is worth noting, however, that such results cannot be generalized due to the short treatment period (i.e., one session and one task). The questions Yanguas (2010) tried to answer, therefore, need to be further examined so that we can obtain more generalizable data.

---

[4]Several turns might elapse before negotiation about particular forms that triggered misunderstanding is fully resolved.

Applying form-focused information-gap tasks via voice chat, Granena (2016) scrutinized whether individual or interactive task performance via voice chat differently impact L2 development. Six groups of participants, a total of 126 EFL learners who were native speakers of Spanish enrolling in an intermediate-level English course online at a Spanish university, took part in this study. The participants were randomly assigned to one of the three target structures (modal verbs, the past tense, and connectors) and were asked to complete a jigsaw task under either an interactive or individual condition. The participants also completed a pre-test and an immediate post-test activity to capture any changes in their performance with regard to the targeted structures. The findings revealed that the EFL learners improved from pre-test to post-test irrespective of the targeted structures and/or whether the learners were working on the task individually or with someone else. Still, when the targeted structure was either the past tense or connectors, learners in the interactive condition outperformed those in the individual condition. Hence the results of this study suggested that collaborative interaction via voice chat also facilitates L2 learning.

More recently, Yangus and Bergin (2018) investigated whether task type and/or mode of SCMC (video vs. audio) have any effects on the number, focus, and outcome of LREs. Seventy-eight intermediate proficiency level Spanish learners (selected from six intact fourth-semester Spanish classes) were randomly placed in dyads, assigned to either video or audio SCMC groups, and worked on two communicative unfocused tasks (i.e., no specific language structures were targeted in the tasks). Repeated-measures ANOVA analyses were performed to identify any significant differences in the number, focus, and outcome of LREs per task and per mode. The results revealed no difference in the number of LREs per task or CMC mode. This indicated that learners do indeed focus on form and produce LREs when interacting in the L2 via audio and video SCMC. However, significant differences were found in LRE foci per task: lexical LREs predominated in the jigsaw task and grammatical LREs in the dictogloss

task. Finally, SCMC mode affected the LRE outcome regardless of the task type; that is, a significantly larger number of unresolved LREs were found in the audio SCMC groups. Hence, this finding could be attributed to the lack of visual cues in the voice chat mode.

It is worth noting that the rationale of choosing voice chat as a tool of interaction in this research is based on Hampel & Hauck's (2004) claim that "recent developments in audio graphic conferencing can now complement written CMC by offering the possibility of going a step further and supporting oral language acquisition as well" (p. 67). Later in this chapter, I will fully address other studies that particularly scrutinized the impact of voice chat on L2 development.

2.2 Components of L2 Proficiency

Various tools of CMC have been incorporated within L2 classrooms and proved to be beneficial for L2 acquisition. The primary focus of recent studies, however, has shifted to address one of the most prevalent questions which lies at the heart of SLA and CMC research: that is, how do L2 learners develop and hence turn into more proficient language users? Skehan (1996, 1998) was the first to devise an L2 model that brought complexity, accuracy, and fluency (commonly referred to as CAF) together to measure learners' proficiency in the target language. This process operates by analysing which features of "the interlanguage system change as acquisition unfolds" and how such change proceeds (Norris & Ortega, 2009: 557). SLA and CMC researchers have incorporated these principal dimensions that constitute the multi-componential nature of L2 proficiency into their research (Abrams, 2003; Ellis & Barkhuizen, 2005; Housen et al., 2012; Hsu, 2017; Michel; 2017; Payne & Whitney, 2002; Robinson, 2011; Skehan, 2009; Skehan & Foster, 2001; Sotillo, 2000). Yet, defining as well as measuring these distinct and competing constructs have been very problematic in SLA research. Housen & Kuiken (2009) claim that when defining CAF, "there is evidence that

agreement cannot be taken for granted and that various definitions and interpretations coexist" (p.3). Housen et al. (2012) further argue that most of the studies that focus on CAF "either do not explicitly define what they mean by these terms, or when they do, they do so in rather general and vague terms…or in terms of concrete psychometric instruments and quantitative metrics" (p.3). This lack of clarity and consensus has, unfortunately, restricted the comparability of findings across previous CAF research and probably explains the contradictory results uncovered in the CAF literature (Housen & Kuiken 2009; Norris & Ortega, 2009).

In the sections that follow, various definitions of the CAF triad will be discussed, highlighting their weaknesses and then arriving at a working definition for each component to fit the aims of the current study.

2.2.1 Complexity

As befits the term, complexity has been extensively recognized for being the most controversial component of the CAF triad with its multidimensional nature (Norris & Ortega, 2009; Pallotti, 2009, 2015). Previous SLA research utilized this term interchangeably to describe either linguistic or cognitive complexity (Thompson, 2014). Cognitive complexity indicates how difficult a particular form/task is to process under various conditions (Housen et al., 2012). For instance, a task that requires high levels of reasoning is considered more complex and more cognitively demanding than one that does not. A further discussion of task complexity will come later in this chapter. Linguistic complexity, on the other hand, as a component on its own represents an array of sub-constructs, including lexical, morphological, syntactic, or even phonological complexity. Thus, even when the term complexity is narrowed down, it is not easy to describe it. The expressions used to define complexity have been seen as being vague and ambiguous (Bulte & Housen, 2012). Some researchers, for instance, believe

that linguistic complexity is "the extent to which learners produce elaborated language" (Ellis & Barkhuizen, 2005: 139). Now, according to Thompson (2014), this ability to produce elaborated language and a variety of syntactic structures might exclude beginner level learners or indicate that it is unlikely they would generate such complex output. In other words, the definition suggests that learners should have sufficient knowledge to produce elaborated forms; knowledge learners typically do not possess at beginner levels. However, this description does not clearly explain what is considered as elaborated language. Skehan & Foster (1999) propose a different definition for linguistic complexity, arguing that it is "the capacity to use more advanced language, with the possibility that such language may not be controlled so effectively" (p.96). In other words, linguistic complexity might illustrate learners' readiness to use more advanced, yet not necessarily well-controlled language. Thus, learners' willingness to produce linguistic features above their current proficiency level (i.e., forms more complex than those they previously acquired) might help them structure and restructure their L2 knowledge, consequently leading to interlanguage development. The main problem with this definition, as suggested by Thompson (2014), is that it is not clear what the researchers mean by the production of 'more advanced language'; that is to say, whether it is associated with their language proficiency or certain grammatical forms.

A more versatile definition of complexity has been proposed recently by Bulte & Housen (2012), suggesting that:

> "A language feature or system of features is seen as complex if it is somehow costly or taxing for language users and learners, particularly in terms of the mental effort or resources that they have to invest in processing or internalizing the feature(s)" (p. 23).

Unlike the definitions stated earlier in this section, this one is limited in its scope. In other words, it narrows down the concept of complexity to mainly refer to particular features of the language—specifically, to features often associated with being cognitively demanding, i.e., too

hard to be used or acquired by L2 learners. Following Thompson (2014), this definition will be adopted in the current research as it closely serves its aims. That is, this study intends to examine learners' use of linguistic forms known for either being difficult to produce or as non-salient in oral speech, especially for EFL learners (e.g., relative clauses, articles, and third -s singular). It is assumed that the SLs will struggle to incorporate these forms into their L2 production since some of them are not part of their L1, or simply because those structures need a mental effort to be internalized.

2.2.2 Accuracy

Whereas complexity seems to capture additional risk-taking attempts on the part of the learners to try out new hypotheses, accuracy illustrates learners' tendency to employ well-controlled structures to achieve a more target-like use of L2. Yuan & Ellis (2003) state that what constitutes accurate production of L2 is "the extent to which the language produced conforms to target language norms" (p.2). In other words, it shows learners' capacity to produce an output that resembles that of NSs and contains a minimum amount of non-target like forms. Similarly, Housen et al. (2012), who view accuracy or what they call 'correctness' as the most transparent construct of CAF, echo Hammerly (1991), Pallotti (2009), and Wolfe-Quintero et al.'s (1998) definition of accuracy. They propose that accuracy primarily refers to the "extent to which an L2 learner's performance (and the L2 system that underlies this performance) deviates from a norm (i.e., usually the native speaker)"; thereby, "deviations from the norm are traditionally labelled *errors*" (p.4).

However, as Pallotti (2009) contends, one must be careful when measuring errors as the speaker "can have perfectly accurate but communicatively inadequate messages (colorless green ideas...)" (p. 592). Such an utterance, although it may grammatically conform to L2 norms, can lead to misunderstandings. Another criticism of Housen & Kuiken's (2009)

definition was later levelled by Thompson (2014), who claims that the term 'deviations' is not fully explained; one cannot be sure whether it refers to grammatical or pronunciation errors. After all, a learner could produce a grammatically correct utterance, yet the way they pronounce it might deviate from the norms of NSs. Responding to previous criticism, Housen et al. (2012) explain that "the A in CAF be interpreted not only as accuracy in the narrowest sense of the term but also as appropriateness and acceptability" (p. 4). That is to say, when formulating an utterance, learners should not only attend to grammatical rules *per se* but should also be careful that the content of their message is comprehensible and adequate, i.e., fits the context in which it is conveyed. Following Thompson (2014), the current study will also adopt Housen et al.'s (2012) definition of accuracy as the main aim is not just to explore learners' correct use of particular forms (e.g., relative clauses), but also to make sure that such use is appropriate, communicatively effective, and comprehensible.

2.2.3 Fluency

Defining fluency as a feature of production has been perceived to be problematic in previous research as well (Housen et al., 2012; Kormos, 2006; Lennon, 2000). The complexity of finding an appropriate explanation to define this phenomenon lies in its multifaceted nature, involving linguistic, psycholinguistic, and sociolinguistic factors (Tavakoli et al., 2016; Thompson, 2014). For some researchers, fluency refers to "the production of language in real-time without undue pausing or hesitation" (Ellis & Barkhuizen, 2005: 139), or "the capacity to use language in real-time, to emphasize meanings, possibly drawing on more lexicalized systems" (Skehan & Foster, 1999: 96). In other words, fluency implies learners' tendency to give more priority to meaning and utilize ready-made chunks to cope with ongoing real-time discourse. These definitions, however, have not gone uncriticised; there has been a lack of consensus concerning whether this pausing is directly related to disfluency or some other

social/personal factors (e.g., anxiety, lack or rapport between the interlocutors). It is also worth noting that in a technology-enhanced learning environment, pausing could be due to some technical reasons, such as connection problems and Internet speed. Another thorny issue concerns learners' ability to balance form and meaning while still producing fluent speech. Lennon (2000), for example, contends that fluency involves "the rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing" (p. 26). In other words, a fluent L2 learner is capable of attending to form and meaning simultaneously. More recently, Tavakoli and Skehan (2005) postulated that fluency consists of three subdimensions: *silence or breakdown fluency*, which refers to the number, length, as well as the duration of (filled/unfilled) pauses: *speed*, which deals with how fast the language is produced (e.g., speech rate, i.e., number of words/syllables per minute): and *repair fluency*, which involves repetitions, false starts, and reformulations of words/phrases. Hence, this definition will be adopted in the current study. This definition that combines different aspects of fluency will be adopted in the current study which aims to examine the effects of different planning conditions on learners' subsequent oral production, where:

"speed is associated with control of and access to procedulised knowledge;

breakdown is thought to reflect the planning and conceptualisation stages of language production;

while repair fluency is seen as an indicator of monitoring processes" (Michel, 2017: 56).

It is worthwhile mentioning that despite the prominence of CAF components as the main variables for measuring L2 proficiency, especially in SLA contexts, most studies tend to omit mention of the definitions they adopt for these components (e.g., Abrams, 2003; Ahmadian,

2012). The sections to follow will synthesise studies that implemented this triad to measure L2 development in both SLA and CMC research.

2.3 Oral Proficiency

Oral proficiency is the individual's ability to produce comprehensible language in terms of syntax, word choice, grammar, and pronunciation (Payne & Whitney, 2002). Lin (2015) also defines oral proficiency as "learners' competence in key traits of oral interactions, including pronunciation, syntactic complexity, lexical complexity, density, richness, overall accuracy and fluency" (p. 266). It has been assumed that online interaction in SCMC serves to enhance oral proficiency (Abrams, 2003; Kost, 2004; Payne & Whitney, 2002). Payne & Whitney (2002) hypothesized that SCMC could indirectly foster oral proficiency by employing the same cognitive processes depicted in F2F spontaneous interactions. In compliance with this hypothesis, Abrams (2003) compared the performance of 96 German intermediate level learners. The participants were divided into three groups: SCMC, asynchronous CMC (ACMC) (e.g., emails), and a control group, and worked on three oral discussion tasks for one semester. Each week, the students in the control group carried out regular class activities F2F, whereas the SCMC and ACMC groups performed these activities orally using online tools. The main difference between CMC groups was in terms of the time they each had to complete the activities: 50 minutes and a week, respectively. All the participants were then invited to take part in a weekly whole-class F2F discussion with the intent to capture any change in their oral proficiency. The assumption was that the learners in the CMC groups would produce an increased number of idea units/words and more instances of lexical diversity, and grammatically correct syntax than the F2F group. General measures were employed to operationalize the syntactic complexity of learners' performance (e.g., C-units and lexical diversity/richness). The findings duly revealed an increase in the quantity of the language

produced in oral discussions by the learners in the SCMC group as compared to the other two groups. Yet the asynchronous CMC group did not outperform the control group in terms of the amount of language they each produced. Due to the lack of recall interviews with the learners, it is hard to single out the causes. Possibly the ACMC mode had a negative influence on learners' motivation due to the extended period of interaction and the fact learners might wait for a relatively long time to hear from their peers. Besides, the data showed no significant differences, either lexically or syntactically, among the three groups. Such an outcome indicates that the quality of the language learners produce online can be very similar to that of F2F discourse. It does not necessarily demonstrate the superiority of online interaction but suggests that this mode is at least as good as traditional oral dialogue.

Similarly, Kost (2004) investigated the effects of different treatments: oral role play, synchronous online discussions, and no treatment on the development of spoken and written proficiency after one semester of instruction. To this end, Kost appropriated a pre-test/post-test design with 94 learners of German. That is, oral interviews and timed in-class writing were employed at the beginning and end of the semester to gauge learners' proficiency levels. The researcher also adopted general and specific measures (e.g., syntactic complexity, subject-verb agreement) as the main variables to assess L2 development. No statistically significant data were found because of the treatments by the end of the semester in terms of learners' oral or written proficiency. However, all groups showed a significant gain in oral and written proficiency due to the instruction. Thus, SCMC sessions did not result in a lower achievement level for the learners in any of the experimental groups.

2.4 CMC and Written Proficiency

As noted earlier, only a handful of studies have examined the effects of text-based SCMC on developing L2 written proficiency. Elola & Oskoz (2010) observed learners' collaborative

synchronous interactions (via wikis and chats) when discussing structure, content, as well as different other factors related to the elaboration of the writing task. Eight Spanish learners enrolled in an advanced writing course participated in the study, where they had to complete two argumentative essays: one collaboratively and the other individually. During the class time (two and a half hour sessions weekly), they worked on grammar exercises and discussed the organization and structure of different written genres. Wikis were used as a platform for both collaborative and individual writing assignments; each task was completed in 15 days as learners submitted the first draft and then a final revised version of their writing based on the instructors' feedback. Text-based and chats were also selected as complementary chatting tools so that the participants could discuss essay content, organization, and form. Elola & Oskoz (2010) adopted general measures to operationalize CAF: the number of words and number of T-units used (fluency), the percentage of error-free T-units (accuracy), and the percentage of words and subordinate clauses per T-unit (complexity). The researchers justified their use of T-units in light of previous research that considered this unit as the most appropriate way to code and record changes amongst different drafts "(Arnold et al., 2009; Polio, 1997; Spelman Miller, 2006)", (cited in Elola & Oskoz, 2010: 56). Although findings revealed no statistically significant differences in terms of CAF when comparing learners' output in the collaborative and individual tasks, there were noticeable trends that illustrate how their interactions with the text vary when working collaboratively or individually. It was clear that with their individual writing, learners tended to focus primarily on the grammatical features of their language. Their collaborative dialogue, on the other hand, showed how they went beyond the confines of grammar to consider the structure and organization of their assignments as well. Putting it another way, they scaffold and help each other to enhance the overall quality of their writing product. A plausible interpretation for the non-significant results could be learners' high proficiency level in Spanish and the fact they were already enrolled in a Spanish course that

discussed different issues related to grammar, organization, and structure. Elola & Oskoz (2010) suggest that "other language proficiency levels should also be considered when conducting research into the application of social tools for FL writing" (p.65). The researchers also presume that the task type could have affected their results; different cognitively demanding tasks might yield varied results.

Bikowski & Vithanage (2016) adopted a pre-test/post-test research model to scrutinize the effects of repeated in-class web-based collaborative activities on the writing development of fifty-nine English learners. The learners were enrolled in a fifteen-week undergraduate writing class. Thirty-two learners worked on four in-class web-based collaborative tasks in addition to extra individual writing tasks (experimental group), and twenty-seven performed the same task types but individually (control group). Based on Elola & Oskoz's (2010) argument stated earlier, Bikowski & Vithanage decided to include different genres for the pre-test and post-tests, as well as for the in-class writing tasks. The findings indicated that both groups achieved significant gains when comparing their pre/post-test scores. However, as opposed to the control group, the experimental group showed significantly more writing gains reflected in their individual writing. These gains might be because the pre-test scores of the experimental group were lower than those of the other group. Accordingly, this supports Elola & Oskoz's (2010) claim and indicates that collaborative online writing may be particularly favourable for lower proficiency level learners. It is worth noting, however, that unlike all the other studies that are reviewed in this chapter, Bikowski & Vithanage chose to analyse any changes in the content, organization, academic style, and grammar of the participants' writing assignments. They did not offer any explanation of what they meant when referring to the gains learners achieved at the end of the study. Such lack of details affects the reliability and generalisability of the results as it is not clear which features of the language had improved due to collaborative online interaction. Furthermore, Presumably, the varied findings of the previous research reviewed so

far could be partially due to the syntactic/morphological measures employed (e.g., T-units, C-units, coordinate and subordinate clauses) to compare learners' performance in text-based and voice chat.

2.5 Factors Affecting L2 Proficiency

L2 proficiency, operationalised by analysing the CAF of learners' production, seems to be affected by different learning conditions, such as task type/complexity, planning time, and learners' anxiety level (e.g., Kuiken & Vedder, 2007; Muñoz, 2006; Skehan & Foster, 1999; Ellis & Yuan, 2004; Yuan & Ellis, 2003). Reflecting on this diverse body of research, Housen & Kuiken (2009) argue that:

> "CAF emerge as distinct components of L2 performance and L2 proficiency which can be separately measured and which may be variably manifested under varying conditions of L2 use, and which may be differentially developed by different types of learners under different learning conditions" (p. 462).

However, the impact certain learning conditions have on the development of L2 proficiency in SCMC is a highly critical area that still needs to be thoroughly explored. Hence, the sections that follow will fully address the importance of including these variables in the current study.

2.5.1 Task Implementation in SLA & SCMC

Numerous explanations have been put forward in an attempt to define the constituents of a good task. The common consensus is that tasks are pedagogical activities that trigger language use with a primary focus on meaning to reach a pre-defined goal (Bygate et al., 2001; Ellis, 2003; Willis, 1996). Two types of tasks have been distinguished accordingly: unfocused vs. focused. Whereas the former does not involve reference to any particular formal aspects of a language, the latter aims to direct learners' attention to process certain linguistic features such

as relative clauses and past tense, making their use *essential* to complete that task (Ellis, 2003; 2010). Ellis' (2003) taxonomy indicates that not all tasks are designed to promote communicative competence and that the execution of the more controlled kind of tasks merely seeks to help students master a grammar point. Hence, this helps explain variations in the results of previous research that utilized different task types/designs to examine learners' performance. Designing a focused task is not as straightforward as it may seem since learners can still find ways (i.e., using communicative strategies) to get around integrating the targeted forms (Loschky & Bley-Vroman, 1993). Samuda & Bygate (2008) subsequently offered a more transparent definition of a task with more details regarding the nature of its outcome. They claim that:

> "A task is a holistic activity which engages language use in order to achieve some non-linguistic outcome while meeting a linguistic challenge, with the overall aim of promoting language learning" (p. 65).

In other words, a task is a whole unit that demands learners' use of their cognitive abilities by employing forms just beyond their current proficiency level. It also requires the learners to use different language skills (e.g., writing and listening) and sub-skills (e.g., vocabulary and pronunciation) to reach a communicative outcome (e.g., solve a problem). Interactive language learning tasks, therefore, have the potential to improve communicative competence. That is to say, during task engagement, learners get the chance to commence, direct, modify, and terminate interactions. Yet real-world cognitive processes are still necessary to achieve that goal and facilitate L2 learning (Chun, 1994; Doughty & Pica, 1986; Gass & Varonis, 1994). As the purpose of this study is to elicit learners' accurate use of particular linguistic features (e.g., articles, relative clauses) while achieving a communicative outcome, I will rely on Samuda & Bygate's (2008) definition and the implementation of focused tasks. Relatively little research on tasks has analysed how variations in the way that tasks are utilized influence

learning opportunities in SLA contexts. Swain & Lapkin (2001), for example, investigated instances of L2 learning during task performance (dictogloss vs. jigsaw)[5]. They predicted that the dictogloss task would provide more opportunities for FoF and language learning compared with the jigsaw, a typical meaning negotiation task. Sixty-two participants were divided into two groups: one performed the jigsaw, and the other worked on a dictogloss activity. The study involved a pre-test, a training session, a mini-lesson (i.e., information on the target forms addressed in the tasks, and a short video to show what learners needed to do during the experiment), and finally, a post-test that contained 'tailor-made' dyad-specific items. Levene's test was conducted for equality of variances. Results revealed that the two tasks used in this study generated fewer differences than what was expected. One plausible reason for Swain & Lapkin's findings might be the mini-lesson that was given before the learners did the task and which in turn served to focus their attention on language form (Yilamz & Granena, 2010). The findings seem to support Ellis's (2010) argument about the effects of giving learners explicit information regarding linguistic forms before task performance. Yilamz & Granena (2010) partially replicated Swain & Lapkin's (2001) study in a SCMC context. They explored whether task type had any effect on the number and the quality of FoF instances triggered during the online interaction. They excluded the mini-lesson integrated by Swain & Lapkin (2001) to eliminate its possible attention-enhancing effect. Five adult intermediate ESL dyads participated in one session of online chat and worked on one jigsaw and one dictogloss task. A total of 25 LREs were identified. The results revealed that the extent to which the dyads were focusing on form was determined by the cognitive demands of different task types. Three dyads did not create any LREs during the jigsaw task, whereas all the dyads did in the dictogloss activity. Besides, a Wilcoxon signed-ranks test showed that the number of LREs produced

---

[5]   The jigsaw involved pairs of students working together to construct a narration based on a series of 8 pictures, 4 pictures to each student. The dictogloss involved students listening to a text and then working in pairs to reconstruct it in writing.

during the dictogloss was significantly higher than that generated by the jigsaw (Z=2.03, p<.05). Hence, different design features and task types reveal different outcomes in terms of L2 acquisition. The findings also indicate that under the pressure of working on cognitively demanding tasks, learners tend to give more attention to language form and hence may produce more complex and accurate output. It is worthwhile examining, however, whether pairing learners with more competent L2 interlocutors, extending the treatment duration, and applying pre-test/post-test (immediate/delayed) design would yield more significant gains regarding L2 development in the long-term.

Due to some drawbacks of text-based interaction, where interlocutors cannot see each other and thereby cannot obtain paralinguistic clues to help evaluate the success of the ongoing interaction, there is a tendency to use abbreviations[6] and emoticons to express the speaker's attitudinal stance (Smith, 2003). It is worth noting, however, that what matters is not just the technology itself but rather how it is being utilized to promote specific learning goals. Likewise, it is crucial to understand how the tasks are being implemented to elicit different versions of learners' L2 production. In other words, designing tasks consistent with the nature of CMC tools might affect the quality or quantity of interaction (Lin, 2015). This theme will be addressed in the following section.

2.5.1.1 Task Design, Task Sequencing, and Task complexity

Previous studies on task complexity have been influenced by two rival theories: one of which is Skehan's (1996, 1998) trade-off effects proposal, whilst the other is Robinson's (2001, 2005) Cognition Hypothesis (Alwi et al., 2012). Both rely on the premise that manipulations of task conditions might affect the way learners' attentional resources are allocated, leading to

---

[6] For instance, 'Brb' instead of saying 'be right back'.

different outcomes in terms of L2 production. In his trade-off theory, for example, Skehan (1996) advocates the concept of the Limited Attentional Capacity Model, which emphasizes the fact that learners' attentional resources are limited and hence attention is restricted to a particular amount of information at a time. He also contends that tasks should be sequenced on a principled basis, so they are not so difficult, "that excessive mental process is required simply to communicate any sort of meaning…" or too "easy that learners are bored…" (p.55). Consequently, Skehan (1996: 52) proposes various criteria for establishing task difficulty as illustrated in **Table 2** below.

Table 2: Skehan's schema for establishing task difficulty

code complexity	syntactic and lexical difficulty
cognitive complexity	content of what is said (conceptualisation)
cognitive processing	the extent to which the learner has to actively think through task content
cognitive familiarity	the extent to which the task relies on ready-made or pre-packaged solutions
communicative stress	Factors which influence the pressure of communication
time pressure	how quickly a task has to be done
modality	speaking/writing, and listening/reading
scale	number of participants, relationships involved
stakes	how important it is to do the task correctly
control	learners' influence on the task; can they ask for clarifications?

Table (2) helps to emphasize a balanced selection of tasks to promote development in CAF. Skehan & Foster (2001) further argue that while various processes (i.e., CAF) compete for attention during cognitively complex task performance, only those that receive enough attention are likely to reach the optimal performance. Skehan & Foster also claim that "tasks which are cognitively demanding in their content are likely to draw attentional resources away from language forms, encouraging learners to avoid more attention-demanding structures in favour of simpler language" (p.189). This assertion seems to contradict the findings of Yilamz

& Granena (2010) and indicates that the nature of oral/written interaction among L2 users is greatly influenced by task type and task complexity. However, Robinson's cognition hypothesis (2001, 2010) illuminates how manipulating certain features of task design and sequencing tasks based on their cognitive demands (i.e., gradually shifting from simple to more complex versions of those tasks) push learners to meet language outcomes in predictable ways to produce more complex, accurate, and fluent output. In his SSARC[7] model, Robinson (2010) claims that sequencing tasks as such allows for 'cumulative learning': while each version presents "an incremental increase in the conceptual and communicative challenge of the task", it stimulates learners to modify and go slightly beyond their interlanguage resources to meet that challenge, generating ideal opportunities for L2 development (p.243). Robinson (2007) believes that each of the CAF components relies on 'multiple attentional pools'; and hence trade-off effects due to increasing task complexity are dubious. Guided by the Cognition Hypothesis, Robinson (2005:5) devised the Triadic Componential Framework for task design, within which he sets out three phases that could affect task performance: task complexity, task conditions, and task difficulty. The framework particularly illustrates how tasks could be sequenced in a way that triggers language production and promotes L2 proficiency. **Table 3** below clearly shows what elements each phase involves.

---

[7] SS: stabilize, simplify
  A: automatize
  RC: restructure, complexify

Table 3: Robinson's (2005) Triadic Componential Framework of Task Complexity

Task Complexity (Cognitive Factors)	Task Conditions (Interactional Factors)	Task Difficulty (Learner Factors)
a. Resource directing +/– Here and now +/– Few elements +/– reasoning demands	a. Participation variables Open/closed One way/two-way Convergent/divergent	a. Affective variables Anxiety Motivation Confidence
b. Resource-dispersing +/– Planning time +/– Single task +/– Prior knowledge	b. Participant variables Same/different gender Familiar/unfamiliar Power/solidarity	b. Ability variables Working Memory Intelligence Aptitude

Apparently, the elements that boost the conceptual and linguistic demands a particular task makes on interaction are all aligned with the task itself (i.e., the language forms learners need to utilize, the given instructions). On the other hand, elements of complexity that add further pressure to access a current language repertoire during task performance are connected to what background knowledge learners already have so that when offered planning time, they could go beyond and extend their existing knowledge. Based on the cognition hypothesis, tasks' manipulation along resource-directing and resource-dispersing variables could make them more or less cognitively demanding. Robinson et al. (2009) hypothesize that manipulating task complexity along resource-dispersing factors; for example, doing a task without any planning time, increases demands on learner cognitive resources yet it does not trigger their attention to focus on formal aspects of the language. Lack of planning time, therefore, could lead to a negative influence on the complexity and accuracy of L2 production. In contrast, increasing task complexity along resource directing variables (i.e., a task with many new elements) might

push learners to produce more complex, accurate forms and incorporate these items into their production, compared with a simple task with only a couple of elements. Fluency, however, could be negatively affected by the processing demands of cognitively demanding tasks due to its nature (Levelt, 1989).

Lee and Polio (2017) argue that "if students keep to writing assignments that elicit simple language, they may not have an opportunity to develop their language" (p. 311). Hence, a substantial body of SLA research has scrutinised how increasing or otherwise decreasing task complexity leads to different outcomes in terms of CAF of learners' production (e.g., Allaw & McDonough, 2019; Ellis & Yuan, 2004; Kormos, 2011; Kuiken & Vedder, 2008; Ong & Zhang, 2010; Robinson, 2001; Skehan & Foster, 1997; Yuan & Ellis, 2003). However, very little is known concerning how different features of a task might impact performance in CMC contexts, and conversely, how the mode of interaction (i.e., F2F vs. online) affects cognitive and psycholinguistic processes. Motivated by Robinson's (2007) theoretical perspective regarding the mediating effects of the cognitive complexity of tasks, Adams et al. (2015) examine whether the learning opportunities and written production of L2 learners vary based on the implementation of different task conditions: the amount of task structure (+/-TS) and inclusion of language support (+/-LS). The first refers to the degree of structural support (i.e., instructions) to be given to learners to perform the task, whereas the second refers to the inclusion or otherwise of (language-focused) pre-task planning that explicitly pushes learners to consider particular language forms. The rationale for including such conditions was to determine their impact on the complexity and accuracy of learners' written production in an online mode. Ninety-six intermediate level ESL learners participated in the study and performed an interactive 45-minute problem-solving task via text chat. General and specific measures were used to operationalise accuracy: mean errors per AS-unit, target-like use of auxiliary verbs, and modal verbs. Both structural (clauses per AS-unit) and lexical complexity

(words/turns) were measured following Norris & Ortega's (2009) recommendation to employ multiple measures of complexity. The results indicate positive effects of enhancing task complexity only in terms of the accuracy of learners' production; learners tend to produce short sentences to avoid non-target like forms. Possibly, as the task increases in its cognitive demands, the learners (given their proficiency level) chose to prioritize accuracy over complexity. One could assume that pairing learners with more advanced L2 users could have helped the learners attend to different aspects of the language simultaneously. Adams et al. (2015) contend, however, that the findings of their study cannot be a basis for generalization due to the short treatment duration; only one session for each group where the participants spent 45 minutes working on the interactive problem-solving task. Thus, longitudinal analysis of learners' interaction might lead to different results.

Révész (2011) also investigates the effects of task complexity on the extent to which learners make form-meaning connections during task performance in a classroom context. Forty-three high-intermediate to advanced level ESL learners, who enrolled in an intensive English course in the USA, worked in small groups during their normally scheduled classes and carried out two versions of the same oral argumentative task (simple and complex). The two versions were manipulated along +/-reasoning demands and the +/few elements dimensions. Thus, the more complex one involved students making more reasoning decisions. To avoid any order effects, the sequence of the tasks was counterbalanced. Each task was performed in three phases. Learners were first given 5 minutes to individually think and make their decision about that task. Then they moved on to discuss the topic in small groups to reach a mutual decision (15-20 minutes). Finally, each small group shared their decision with the whole class group (10-15 minutes). Both global (e.g., number of syntactic clauses per AS-unit, ratio of errors to AS-unit) and specific (e.g., use of conjoined clauses) measures were analysed in order to ascertain the effects of task complexity on language learning opportunities. Results of quantitative analyses

indicated that more complex versions of the task have a positive impact on prompting the use of advanced clause types (e.g., if, when) and that more accuracy and lexical diversity of speech production resulted when task complexity was enhanced. These findings seem indeed more compatible with the Cognition Hypothesis than those of Adams et al. (2015). However, the findings also revealed that the overall syntactic complexity of learners' production decreased on the more complex version of the task. According to Révész (2011), this is in line with Robinson (2005), who predicts "lower structural complexity on more complex interactive tasks owing to the amplified amount of negotiation that tends to result when the cognitive demands of interactive task are increased" (p.176). Hence, perhaps this decrease in structural complexity was due to the nature of turn-taking and some interactional moves (e.g., clarification requests, comprehension checks) that prevent learners from building complex linguistic structures.

Several studies have examined the effects of the planning condition as a resource-dispersing variable in F2F settings (e.g., Ahmadian, 2012; Ellis & Yuan, 2004; Foster & Skehan, 1996; Ong & Zhang, 2010; Skehan & Foster, 1997). However, only a handful of studies have examined this in different modes of SCMC environment (e.g., Hsu, 2017; Satar & Ozdener, 2008). Some of these studies will be fully discussed in the section below.

2.5.2 Task Planning

Task implementation conditions, especially those involving manipulations of planning conditions, have been found advantageous to direct learners' attentional resources to particular forms and positively affect their L2 performance (see Ellis, 2009 for a review). The following sections, therefore, will give more details regarding this variable and the different planning conditions that have been employed in previous research.

2.5.2.1 Planning Conditions in SLA

A growing body of SLA research, conducted within the framework of task-based language learning, is built on the premise that the quantity and quality of L2 production are likely to be enhanced under planned conditions (e.g., Ellis, 2003; Foster & Skehan, 1996; Ortega, 1999; Schmidt, 2001; Williams, 2005). Such an implementation variable seems to lead to a relatively consistent impact on various aspects of production, including the content learners communicate when carrying out a particular task and their own choice of the language. Thus, planning could be a straightforward means to control language production and indirectly mediate interlanguage development. Accordingly, it might be a pedagogically beneficial tool for language teachers as well. Given that SLs in the current study are not used to L2 interaction, it would be interesting to investigate whether allowing planning opportunities might conceivably ease the burden on their working memory and results in more developed production.

The Cognition Hypothesis (Robinson, 2001) has motivated the research on task planning, drawing on several theoretical constructs. Information processing theory (Skehan, 1998), for example, highlights the limitations of learners' working memory in attending to form and content simultaneously. This theory has formed the cornerstone for all the studies in this area. Researchers postulate that learners might feel obliged to favour one aspect of the language over the other; in most cases, when attentional resources are under pressure, there is a tendency for learners to prioritize content at the expense of form. However, Robinson (2005) contends that planning and other possible task implementation conditions (e.g., task type, task structure) might facilitate balanced attention to the CAF triad. In the same vein, Ellis (2005) claims that affording learners the chance to plan may ease the burden on their Working Memory, making it possible for them "to engage in controlled processing and to process multiple systems linearly" (p.8). Thereby, planning creates an environment where learners could map form onto

meaning via integrating recently learned linguistic forms into their L2 production. Hsu (2017) proposed that:

> "The provision of planning time can therefore serve as a means to help learners compensate for these processing limitations and divert their attention to linguistic form during task production in a self-regulated manner" (p. 359).

The effects of planning on L2 production, believed to occur due to offering learners time to plan their task performance, have frequently been discussed in the light of Levelt's (1989) model of speech production (e.g., Ahmadian, 2012; Hsu, 2017). Levelt (1989) views "the speaker as a highly complex information processor who can, in some rather mysterious way, transform intentions, thoughts, feelings into fluently articulated speech" (p. 1). His model is one of the most influential psycholinguistic perspectives that was originally devised to describe first language production. Yet it has then extensively been used in previous L2 research to analyse how language is produced (e.g., Ahmadian, 2012; de Bot, 1992; Hsu, 2017; Payne & Whitney, 2002). As shown in Figure 1 below, this model comprises three stages of language production. It begins with the 'Conceptualizer,' where the message content is generated. Once the message is maintained in the Working Memory, it is then upgraded to the following stage, 'the Formulator,' where grammatical and phonological encoding take place, determining the surface structure and resulting in an articulatory plan of the intended utterance. It is worth mentioning that in this stage, the speaker still has the chance to edit and monitor the message with the help of subvocalization before it is stored in the so-called "Articulatory Buffer (Working Memory)" (Payne & Whitney, 2002: 10). This plan is directed afterwards to the last phase, the Articulator, where it becomes ready to be produced. Levelt's model then seems to involve competition for limited attentional resources. Accordingly, there might be a positive impact of planning time offered in the SCMC mode on L2 proficiency: the features of text-based chat in particular (i.e., on-screen messages) could reduce the cognitive demands for

language production on the Working Memory. Therefore, there would be a higher tendency for L2 learners to process the incoming input smoothly and produce more complex output compared with F2F contexts. Hence Levelt's model is employed in the current study to explore the interaction between different modes of SCMC, strategic pre-planning/no planning conditions, and L2 production.

Figure 1: Levelt's (1989) model of speech production (Payne & Whitney, 2002: 11)

Figure 1
Blueprint for the Speaker
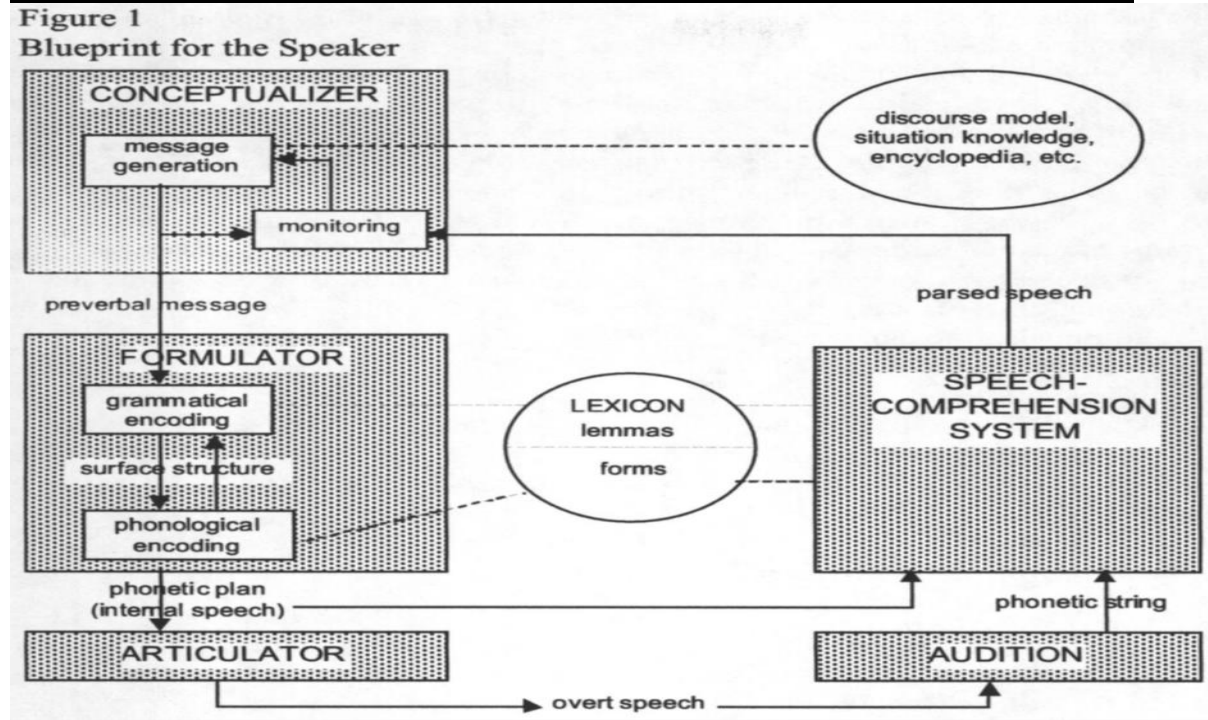
Ellis (2005) divides task planning into two categories: pre-task planning and within-task planning (see **Figure 2**). Each has been further split into other types, which will be addressed fully in the following subsections. According to Ellis (2005), each planning condition has a uniquely positive impact on different aspects of performance[8]. Therefore, a setting in which pre-task and within-task planning are both feasible would create an ideal context for learners to maximize their interlanguage development.

---

[8] Pre-task planning seems to have a positive impact on complexity and fluency, whereas within-task planning is found more beneficial for enhancing accuracy.

Figure 2: Task planning (adopted from Ellis, 2005: 4)


```
                                           ┌─── Rehearsal
                         Pre-task planning ─┤
Planning ────┬───────────                   └─── Strategic planning

             └───────────                   ┌─── Pressured
                         Within task planning┤
                                            └─── Unpressured
```
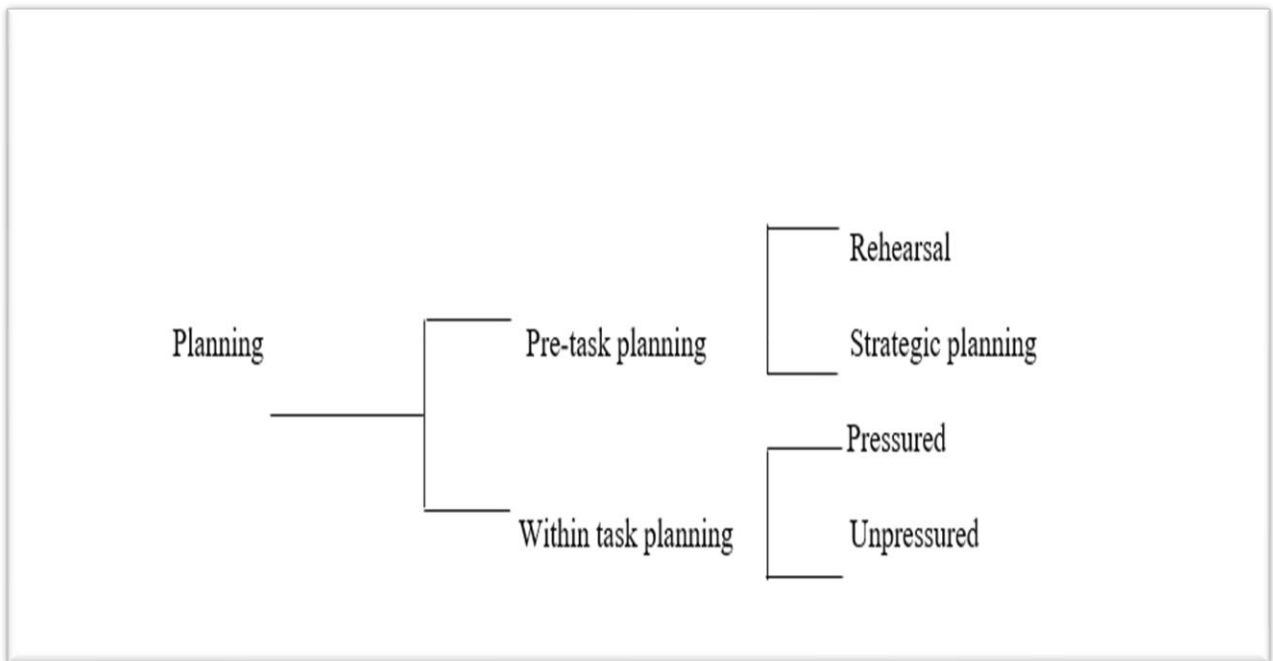
2.5.2.1.1 Pre-task planning: strategic vs. rehearsal

The assumption underlying pre-task planning is that learners have some time to regulate their performance, orient themselves to the demands of the task, and access their interlanguage grammar (Ellis, 2005). Consequently, a broader linguistic repertoire would be available for their subsequent online use (Crookes, 1989). There is some disagreement about how planning prior to task performance affects attention. Foster & Skehan (1996), for example, contend that language learners are likely to produce more complex and fluent language when offered a chance to pre-plan their tasks than when they are not. An alternate view, propagated by Robinson (2005), is that planning a task in advance serves to simplify that task, leading to greater accuracy as it facilitates automatic access to the stored language. Two types of pre-task planning have been identified:

a. Rehearsal planning[9], known as task repetition, is widely investigated in SLA research and entails giving learners the chance to carry out the task before the actual performance (Ellis, 2005). Previous findings revealed a positive impact of this form of planning on learners' subsequent L2 performance of the same language task, especially regarding complexity and accuracy (Ahmadian & Tavakoli, 2010). However, there is hardly any evidence that this impact is transferable to another language task, even if the latter was of the same type as the original one.

b. Strategic planning involves providing learners time to prepare for a particular task, consider the content they need to encode, and finally, seek ways to express that content (Ellis, 2005). This type of planning could be either guided, that is, learners are given step-by-step instructions as to how to plan, whether they need to attend to specific linguistic forms, to meaning, or maybe both; or unguided, which entails leaving learners to their own devices with no guidance provided so that they could freely plan the task at hand. The guided planning condition could be beneficial for triggering learners to use hard-to-learn structures (e.g., relative clauses) or non-salient (e.g., articles) linguistic forms during task performance, thus aligning attentional mechanisms to specific aspects of the language and facilitate L2 development.

Mochizuki & Ortega (2008) claim that strategic planning, which primarily refers to grammatical forms, leads to significant improvements in terms of learners' accuracy. However, previous researchers claim that the effect of strategic planning on both complexity and fluency is greater than that on accuracy: learners tend to draw a conceptual plan of what they would like to say rather than being concerned about framing the linguistic details of their utterances, which in turn, seem hard to carry over into the actual task performance (Ellis, 2005; Kawauchi,

---

[9] This type is only briefly explained as it is beyond the focus of the current study.

2005; Wendel, 1997; Seyyedi et al., 2013; Skehan, 1998; Yuan & Ellis, 2003). Thus, complexity and fluency are often favoured at the expense of accuracy due to trade-off effects. Comparing the impacts of guided, unguided, and no planning conditions on three different types of tasks (personal, narrative, and decision making[10]), Foster & Skehan (1996) reported varied results in terms of CAF. Overall, planners paused less frequently than non-planners; nevertheless, those grouped under the guided planning condition and instructed to focus on lexis, syntax, content, and organization of their production were notably more fluent when performing the narratives than the other two tasks. In contrast, when considering how accurate learners' production was, mixed effects were observed due to the planning conditions. Unguided planning resulted in greater accuracy during the personal and narrative tasks yet not in the decision-making one, which was considered more cognitively demanding than the others. Possibly this explains why learners were less accurate when not given guided planning compared with the other tasks. These findings show that the impacts planning has on accuracy are not clear-cut; that is, they might be affected by other factors (i.e., task type in this case). In terms of the complexity of learners' production, planners were also found to produce more subordination compared to non-planners. This result was found to be statistically significant and was true for all the task types. However, when replicating the study, Skehan & Foster's (1997) findings showed a trade-off effect between accuracy and complexity: planners' production, compared with that of the other tasks, was more accurate but less complex during the narrative. In sum, the results of Foster & Skehan (1996) have rendered empirical support for the claim on how planning conditions interact with task type and differently affect the three dimensions of production.

---

[10] **Personal task:** the participant had to describe to their interactant how to get to their home from college and then to turn off a gas cooker that was left on.
Narrative task: the participants had to build a storyline from a set of pictures that were loosely but not obviously connected.
Decision-making task: the participants had to agree with their partner on a prison sentence for a list of offenders.

Further research is still vital to pinpoint the ways planning interacts with task implementation and mode of interaction, whether and to what extent the SCMC format produces consistent results with these non-CMC studies. At the time of writing, none of the previous studies addressed the delayed effects (i.e., gains over time) of strategic planning on L2 development, neither in an EFL context in general, or in an online mode in particular.

2.5.2.1.2 Within-task planning

This type of planning relates more to the time afforded to learners while performing the task to help them prepare what they want to say and how to. Although within-task planning is a variable beyond the focus of the current study, this brief review is included as it is anticipated that the text-based mode would give learners some time to contemplate their utterances, hence resulting in more accurate L2 production. Hsu (2012) examined the joint effects of pre-task and online planning (i.e., within-task planning) on the complexity, accuracy, and fluency of learners' production in a written SCMC environment. Thirty ESL learners, with intermediate English proficiency, worked on a picture-based task and narrated their version of the story online to the researcher. The learners worked under two groups; only one of the groups was given ten minutes to pre-plan their performance. Both groups, however, were offered unlimited time to complete the task. A combination of CAF measures was used to gauge learners' production, including ratio of clause per AS-unit, lexical variety, percentage of error-free clauses, accurate use of grammatical verb forms, and number of disfluencies. The findings revealed that the opportunity to pre-plan prior to the actual task performance did not lead to more complex, accurate, or fluent language production when unlimited online planning time was available. It was assumed that the unlimited online planning enabled the learners to plan the conceptual content as well as the formulation of their messages. Hsu (2012), therefore, concluded that "the effect of online planning is strong enough to offset the effect of strategic

planning since the learners could use the additional online planning time afforded by the very nature of text chat to both formulate story content and encode it with appropriate linguistic forms even in the absence of strategic planning" (p. 630).

One can notice the absence of research on within-task planning, a mental activity which is problematic for researchers to observe as evidence of its presence is highly unlikely. However, the rise of screen-capture technology in SCMC studies has encouraged research into within-task planning generally, one example of which is Smith's (2008) study that examined learners' self-repair during text-based interaction. Smith tried to test the hypothesis that online planning available via text chat might benefit the complexity and accuracy of L2 performance as learners frequently try to attend to their production. Using Camtasia screen capture software, Smith captured significantly more self-repair than that revealed by analysing unenhanced chat logs alone. In the latter, the researcher can only view the learners' message after it has been sent, in which case, many interesting segments, such as deletions and modifications made before sending the message, will be missed. Sauro & Smith (2010) also found that learners' production that entails evidence of within-task planning (measured as instances of self-correction and deletion when typing one's messages) was syntactically more complex than production that does not.

2.5.2.2 Investigating Planning Conditions in SCMC Research

It has been assumed that planning time afforded during the online mode (text-based chat, in particular) may grant learners better chances to balance attention on the three aspects of the CAF triad. Only a handful of studies, however, have examined the efficacy of task planning on L2 performance and L2 development in text-based SCMC (Hsu, 2017; Sauro & Smith, 2010; Ziegler, 2018). Hsu (2017) recruited thirty-three intermediate to advanced level ESL learners

to examine the combined effects of rehearsal[11] coupled with careful online planning (ROP) and careful online planning alone (OP) on L2 development in text-based SCMC. The participants performed a set of four picture-based narrative tasks (where they needed to describe their picture to an online partner) under the ROP and OP conditions via MSN for two consecutive weeks. In the first week, two of the tasks were utilized as experimental tasks to measure the immediate impact of planning on complexity and accuracy. Under the ROP condition, ten minutes were given to the learners to rehearse the task, during which they typed their description of the pictures using Microsoft Word. This document was saved and closed as learners were not allowed to use it during the real performance. Afterwards, the participants interacted online (via MSN) with the researcher and worked on the same task under no time pressure. The OP group, on the other hand, was not given any time to rehearse before the actual performance. The other two tasks were employed as post-tests to define the subsequent effects on L2 development in the following week. Most of the measures used to gauge learners' production were the same as those used in previous planning research (e.g., mean number of clauses per AS-unit, subject-verb agreement, tense, percentage of error-free AS-units). Immediate performance of the tasks revealed positive effects of both types of planning conditions on complexity; however, the ROP condition resulted in a much more accurate integration of grammatical verb forms. The post-test revealed that the ROP condition resulted in improvements in terms of clausal complexity. The learners also showed more control of correct grammatical verb forms and avoidance of general errors.

Ziegler (2018) also examined how "different pre-task planning times (no planning time, one minute of planning time, and three minutes of planning time" affected learners' production (p. 196). Forty-four intermediate learners of English with B2 proficiency level, and who were

---

[11] As mentioned earlier, rehearsal planning is a condition where the learner has the chance to work on the same task twice; the first is considered the rehearsal planning stage, whereas the second is counted as the real performance of the task.

enrolled in an intensive language learning program, took part in this study. The participants were asked to work on three picture-narrative tasks and collaborate with their L2 partners to create a narration based on these pictures via means of text-chat. Learners' text-chat production was operationalised using a range of general CAF measures, including number of clauses per AS unit, number of words per clause, lexical complexity, percentage of error-free clauses, and number of dysfluencies produced during each task. Results revealed that compared with the other conditions, only the three minutes of planning time yielded increases with regard to lexical complexity; no significant findings were identified though for syntactic/phrasal complexity, accuracy, or fluency. Perhaps the short planning time the learners were offered to prepare for their online performance, the lack of guided instructions in terms of what aspects of the language the learners needed to focus on, as well as the task type (which was meaning-oriented by nature) could have led to the non-significant findings for the rest of the CAF measures.

Given the bulk of research that has been done to examine the influence of planning conditions, especially in an SLA context, to date, very few studies have sufficiently targeted planning as a process (e.g., Ortega, 1999; Sanguran, 2005); and those that have are all non-CMC studies. That is to say, very little is known as to what learners really do when asked to plan, the strategies they follow, and whether they manage to do what they planned to do during the actual performance. The significance of such details is twofold: they afford evidence about whether learners plan as intended and serve as a basis for drawing up guidelines for the design of efficient future planning instructions. Understanding that planning opportunities help learners produce high-quality discourse, Ortega (1999) inspected the strategies learners use when engaged in pre-task planning in a F2F environment, hoping to gain an insight into the cognitive processes they go through while planning. She also interviewed them about their perceptions of the planning opportunities they had in advance of the actual task performance. Presumably,

this would help researchers and teachers understand how planning works and why it might not be suitable for every learner, at least not to the same degree. To this end, she employed retrospective interviews, which revealed that learners adopted an identifiable approach when planning, prioritizing content and organization of ideas before making a conscious effort to attend to any linguistic forms. The measures Ortega (1999) investigated could be the reason for the negative effects on accuracy. She examined learners' accuracy regarding their article use: forms that have been frequently considered low in their saliency and difficult to notice. Based on the gaps identified in this section, much work is vital to examine the effects of planning on L2 performance, especially in a SCMC context. Due to the time/word limit, the present study will only explore one type of planning in full: strategic pre-task planning. The effects of learner factors on L2 development will now be discussed in the following paragraphs.

2.5.3 Learner factors: anxiety levels

In their affective filter theory, Krashen and Terrell (1983) argue that the process of learning/acquiring another language is more effective if learners are emotionally stable and have low affective filters (i.e., low levels of anxiety) (cited in Thompson, 2014). CMC tools seem to provide a liberating experience where learners feel less concerned and less anxious about making errors. Adopting a new approach to analyse L2 performance in SCMC, Satar & Ozdener (2008) investigated whether the mode of interaction (i.e., text-based vs. voice chat) influences verbal skills and anxiety levels. Such an endeavour ties mode and anxiety with learners' output. The pre-test/post-test design entailed analysing data elicited from two experimental groups (text and voice); each consists of thirty EFL learners whose proficiency levels ranged from beginner to pre-intermediate. They also included a control group, which was not allocated any out-of-class activities. Satar & Ozdener (2008) hypothesized that language anxiety levels in both experimental groups would decrease by the end of the study,

whereas those of the control group would remain the same. The experimental groups worked online on eight tasks (two per session) that differed in type (e.g., problem-solving, jigsaw, decision making, and information-gap) over a month; one session of 40-45 minutes per week for each dyad. The participants worked in pairs, and the same dyads were maintained throughout the study. Pre-anxiety/post-anxiety questionnaires and pre-/post-speaking tests were employed for data collection. The results of these tests showed an increase in the speaking proficiency for both groups (i.e., text-based and voice chat) and a reduction in the anxiety levels for the text chat group only. Satar & Ozdener suggested that this could be due to learners' concerns about pronunciation and comprehension. Perhaps this could also be traced back to learners' low proficiency level and the fact that one month was not enough for them to get used to the spoken modality. As voice chat entails speaking skills, it puts more pressure on learners' cognitive processing than the text-based mode. Therefore, one could assume that scaffolding learners to move gradually from one modality of CMC to another would result in more significant results in terms of learners' overall performance and their language anxiety levels. According to Satar & Ozdener, the findings provide evidence that online interaction can aid L2 proficiency when implementing well-designed tasks. However, it seems reasonable to suggest that the short treatment period could have affected the data reliability. Longer-term effects of online interaction (text-based/voice chat) have not been investigated adequately and thoroughly before we can be confident about the effectiveness of adopting the modality as a common practice for enhancing L2 proficiency. Therefore, the current study will expand on this line of research and follow a longitudinal design to measure language anxiety of intermediate level learners as they are interacting with NSs and more competent L2 users. However, understanding the pressure learners might experience due to that interaction, they will be asked to work on the text-based mode for six weeks and then move to voice chat. Such a

learning/teaching environment will probably result in lower anxiety levels and better quantity/quality of L2 production.

2.6 Perceptions and Attitudes towards SCMC

Among the body of research that has investigated SCMC and its effectiveness for language learning and language development, very few studies also refer to learners' attitudes towards this online learning experience (Canals, 2020; Lai & Zhao, 2006; Satar & Ozdener, 2008; Warschauer, 1996; Zeng & Takatuska, 2009). Lai & Zhao (2006) examined ESL learners' perceptions of two conversation modes (i.e., F2F and SCMC) in promoting their noticing of erroneous patterns and the feedback they received from their partners. Learners reported that in SCMC, they had more time to review their output and attend to their partners' feedback. Besides, they indicated that the visual saliency of learners' utterances helped them notice their errors and make necessary revisions. In Warschauer's study (1996), ESL learners had slightly better attitudes on average towards the electronic discussion than F2F communication. They frequently touched upon the advantages online discussions offer; learners were, reportedly, able to express themselves more comfortably and creatively than they did when interacting F2F, and so presumably learners found online discussions assisted their thinking ability. Learners' feeling more comfortable during the online interaction was associated with the thinking time they had to process information compared with F2F conversation.

Additionally, Zeng & Takatuska (2009) designed a survey to elicit EFL learners' perspectives on their online collaborative learning experience. The responses were very positive: "they thought they enjoyed the collaborative learning process, were willing to offer help or accept the suggested solution when faced with a language problem and could make a joint effort to carry out each task" (p. 441). When asked about their online experience and how anxious they were in the text-based and audio chat environments, learners in Satar & Ozdener's (2008) study

stated that they felt less anxious when doing the sessions in pairs. Twenty percent of the voice chat group mentioned that they were less concerned about pronunciation than normal. The participants also reported that they would have been worried about understanding their partner if they had been interacting with a NS rather than another NNL, whom they already knew.

As the literature review presented here suggests, no study so far has examined the combined effects of manipulating task complexity along with pre-planning conditions on L2 performance in two different modes of SCMC (text-based and voice chat). Hence, in light of the Cognition Hypothesis and the Triadic Componential Framework, the current study aims to bridge this gap with the aim of helping inform the design of tasks for online language teaching/learning contexts and providing evidence as to the effectiveness of different SCMC modalities for promoting L2 proficiency. Besides, given the lack of research on learners/teachers' attitudes and perceptions towards online learning/teaching and as highlighted in the study research questions below, the current study will build on previous studies and analyse participants' attitudes/perceptions towards the online experience. The qualitative part of this research also aims to address other under-researched themes in the literature: the planning strategies learners follow when allowed to pre-plan a task and their awareness about any changes in their L2 production due to the online mode of interaction.

2.7 Aims and research questions

Given the ubiquitous use of technology in L2 teaching/learning, it is pertinent to examine how the oral and written modalities of SCMC could boost L2 proficiency. Ultimately, this study was designed to investigate whether the longitudinal interaction with more proficient language users via text-based and voice chat promote L2 development. In particular, the primary aims that guided the conception and design of this thesis were as follows:

- To examine the combined effects of manipulating task complexity along with pre-planning condition in two different modes of SCMC on L2 proficiency.

- To detect any short-term and long-term gains in learners' written and oral proficiency due to the longitudinal online treatment with more proficient L2 users.

- To identify the planning strategies adopted by the SLs prior to as well as during the online interaction.

- To highlight the factors that have impacted learners' L2 development or otherwise.

- To elicit learners'/teachers' attitudes and perceptions of the online experience as a whole.

These aims were explored through the following research questions:

1. Does manipulating task complexity during online sessions have an impact on learners' L2 development?

2. Does the oral and written proficiency of intermediate level Syrian learners improve as a result of their longitudinal online interaction with more proficient L2 interlocutors?

3. What strategies did the Syrian learners use when planning for different task types across different modes of online interaction (text vs voice chat)?

4. What factors, if any, impacted learners' development across different modes of online interaction (text vs voice chat)?

5. What are the attitudes and perceptions of the participants towards the online experience?

3.   <u>Methodology</u>

This chapter describes the design of the current research and demonstrates how it was revised and modified based on the results of the piloting phase. The first two sections give information about the preliminary design of the project and the instruments used to conduct the pilot study. Then the following two sections provide justification of the research strategy (i.e., the longitudinal research design and the mixed-methods approach). This chapter also outlines details of the participants and the materials they used during the main study. Finally, the last two sections describe the procedures according to which the data were collected, coded, and analysed.

3.1 The Initial Design of the Study

In this section, I will explain how the research design was envisaged and why it was originally conceived as such. However, later (in **section 3.5**), I will detail the reasons why this design seemed impractical, given various constraints highlighted during the piloting phase, and how it was eventually modified as a result of these difficulties.

The preparatory plan (i.e., the original design) of the main project involved two separate online studies (one to be conducted via means of text-based chat and the other to be done in a voice chat environment), each incorporating a pre-test/post-test design, and aimed to recruit thirty intermediate-level SLs by randomly pairing them with seven English STs and eight NSs (see **figures 3 & 4** below). All the pairs needed to perform a thirty-minute online session every week and work on a sequence of oral and written tasks that varied in type and increased in complexity over the duration of twelve weeks. SLs would also be required to experience different treatments during the online sessions and operate under one of the following planning/no-planning conditions: the guided pre-planning group, where learners would have ten minutes to prepare for the given task and receive instructions that helped them attend to

particular aspects of grammar and vocabulary (e.g., relative clauses); the unguided pre-planning group, where learners would get the same amount of time to pre-plan their online performance, yet would not receive any explicit instructions to focus on specific language forms; or the control group, where the learners would have no time to plan and rather be asked to work on the online task with their partner immediately. Thus, complexity would be manipulated via means of task type/topic and planning conditions. Each online session would be followed by a thirty-minute interview with the learner and a weekly report written by the learners to freely reflect on their L2 development. Interviews were also to be conducted with all the STs and NSs at the end of the study to enquire about their perceptions regarding the whole online experience.

Figure 3: Study 1 (Text chat mode)
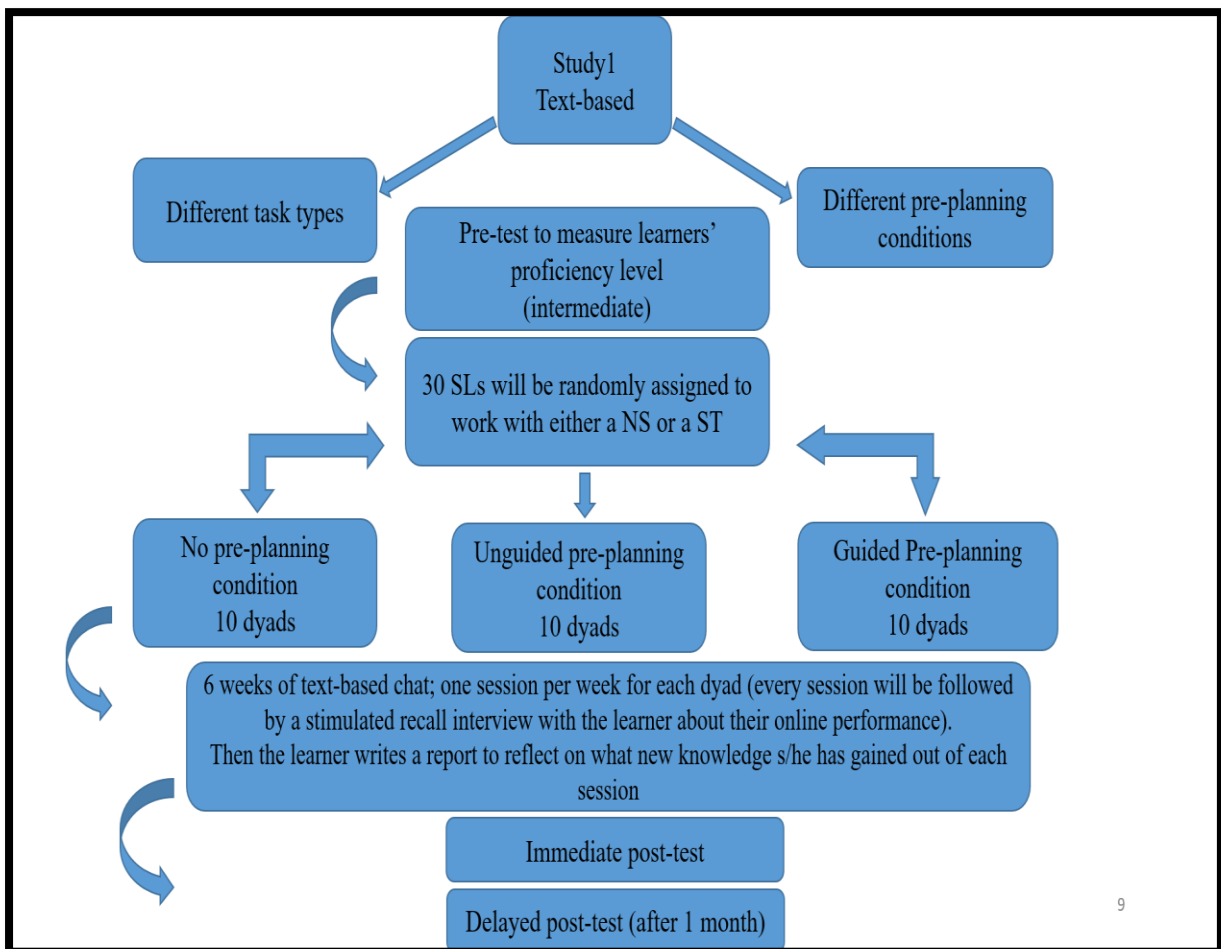
Figure 4: Study 2 (Audio chat mode)
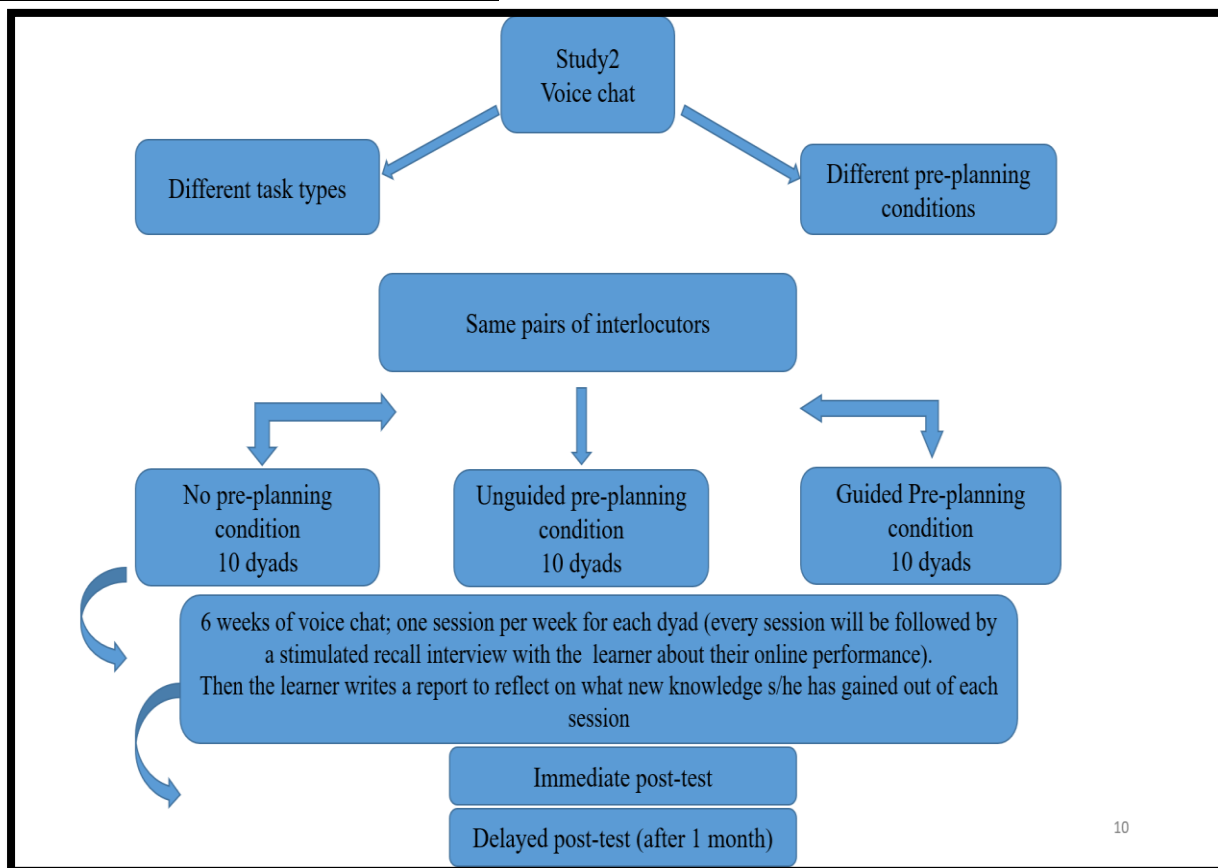
It was assumed that this design would address all the gaps identified previously in the literature review chapter and provide a robust study, which aimed to investigate the combined effects of manipulating task complexity along with planning conditions via SCMC on L2 development in a foreign language context. After obtaining University ethics approval on the abovementioned design, I started collecting data for the piloting phase. **Figures 3 and 4** illustrate each stage of the initial research design.

The primary plan was to quantitatively analyse learners' written and oral production during the online sessions as well as the post-tests based on a set of CAF measures (e.g., number of AS-units, number of clauses, number of errors, number of pauses). On the other hand, learners' planning strategies and the participants' perceptions towards the online experience would be examined via means of stimulated recall interviews and weekly written reflective reports and then qualitatively analysed. Unfortunately, several restraints related to participants' availability

and time/word limits made this research plan unfeasible, and hence a more realistic and modified version of this design was put in place. Thus, the modifications made to the original design were as follows:

- ❖ Instead of conducting two studies, one longitudinal study was carried out via two modes of SCMC (text chat and voice chat)

- ❖ The participants were divided into two groups only: guided planning group (or treatment group), and no planning group (which acted as a control group). That is, one of the planning conditions (i.e., the unguided planning group) was eliminated.

- ❖ Accordingly, the number of the participants was decreased to twenty SLs (instead of thirty), five NSs, and six English STs.

- ❖ Learners' performance throughout the online sessions was excluded from the quantitative analysis, and thus only learners' production during the pre/post-tests was analysed quantitatively.

To avoid repetition, all these amendments will be fully discussed in **section 3.3.3**. It is worth mentioning that the research questions were also revised as follows:

1. Does manipulating task complexity during online sessions have an impact on learners' L2 development?

2. Does the oral and written proficiency of intermediate level Syrian learners improve as a result of their longitudinal online interaction with more proficient L2 interlocutors?

3. What strategies did the Syrian learners use when planning for different task types across different modes of online interaction (text vs voice chat)?

4. What factors, if any, impacted learners' development across different modes of online interaction (text vs voice chat)?

5. What are the attitudes and perceptions of the participants towards the online experience?

3.2 The Pilot Study

This was conceived as a small-scale replica of the main study; a process that has been firmly advocated by many researchers (e.g., Mackey & Gass, 2015), in an attempt to test the research instruments/procedures and obtain preliminary answers to the research questions (which were slightly modified in the main study, see **section 3.1**). Different tasks, planning conditions, instructions, and CAF measures were piloted in order to gauge task difficulty, fine-tune and develop aspects of the procedures. Robson contends that this stage is essential in the research process since it "helps to throw up some of the inevitable problems of converting your design into reality" (2002: 383). Hence, the experience gained from this phase would help me decide on what vital amendments were required to better improve the design of the main project, as well as scrutinise the appropriacy/validity of the implemented materials and instruments. The following sections succinctly address the main components of the pilot study, the procedures followed for data collection, and outline the materials/apparatus trialled at this stage.

Six SLs, two STs, and one NS[12] were randomly put into six dyads and divided into three different planning/no planning groups (two dyads within each group). Following a pre-test that measured learners' proficiency in terms of particular CAF measures (see **section 3.3.8.1.1** for a full description of these measures), the dyads worked on different task types via text-based chat for three weeks. Then the SLs took an immediate post-test as well as a delayed post-test (two weeks later). A similar routine was followed for the voice chat sessions. Stimulated recall interviews were also conducted with the pilotees following every session to ask about their performance and attitudes towards the tasks, instructions, planning conditions, mode, etc. All the interviews were carried out in Arabic with the pilotees in order to lessen the cognitive load they might encounter while recalling their thoughts.

---

[12] A representative sample of the participants to be included in the main study in terms of profile and proficiency (see **appendix H for pilotees' background information**).

The pilot study revealed a number of implications for the main study: first, some learners mentioned that they were struggling to open up during the online interaction as they felt intimidated by the language proficiency of their partners and the fact that they were talking to complete strangers. Therefore, the participants were asked to use the first three/four minutes of the first chatting session to introduce themselves, to get to know more about each other, break the ice, and build some sort of rapport with their partner (Smith, 2003). Second, as different task instructions/types[13] were tested, the following problems were identified and solved accordingly:

3.2.1 Narrative Tasks:

The participants were supposed to view a series of pictures (see **Figure 5**) and attempt to create a story describing their content. Despite the fact that the task was adopted from previous research (Mochizuki & Ortega, 2008: Thompson, 2014) and never reported to cause any problems, participants in the current study found it very disorienting. One of the learners commented:

> "I really got confused when I noticed different people in the pictures; thought it's just the mother and her son and daughter, but then realized there were other characters in the pictures as well, so I wasn't sure who is who."

---

[13] See appendix for all task types used in the piloting phase.

Figure 5: narrative task 1
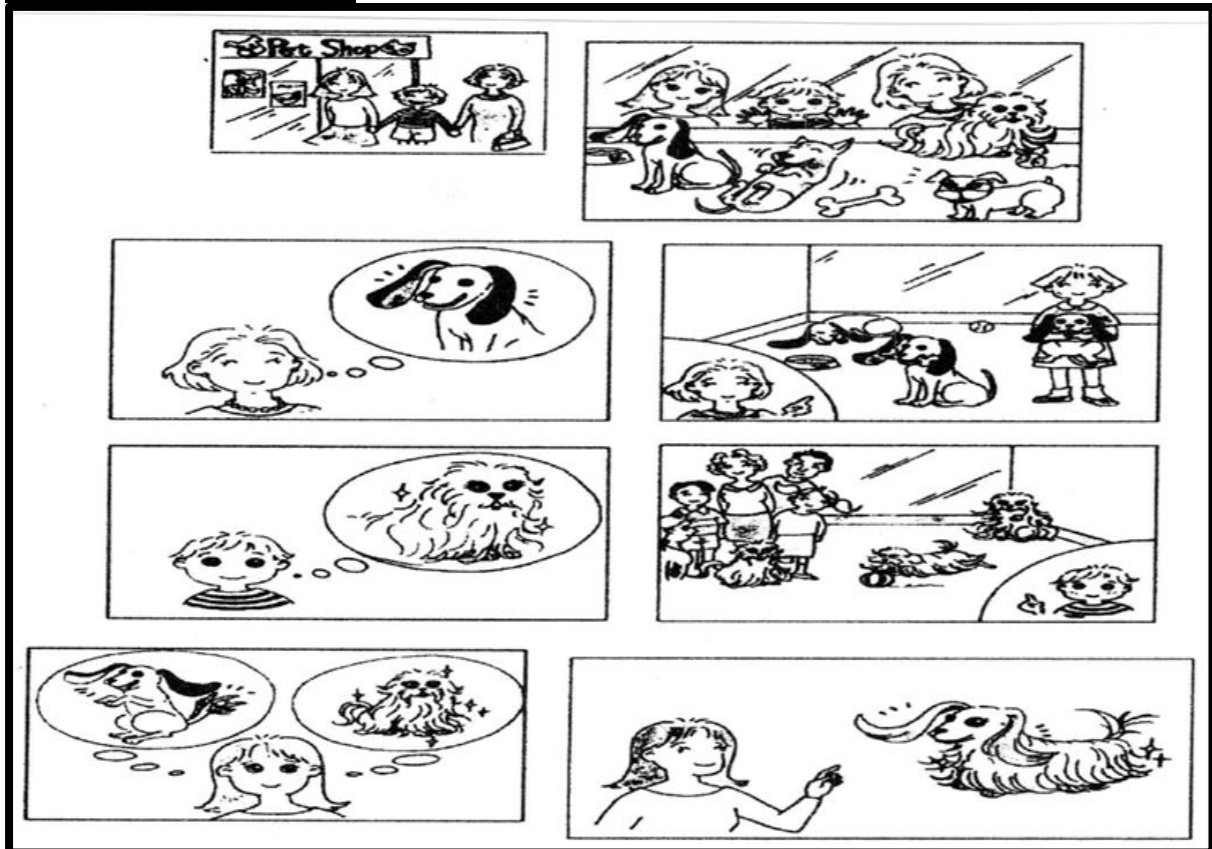
A second picture story was piloted with a different pilotee but also fell foul of the same problems as the first; learners' perceptions regarding these tasks and their appropriateness were almost identical. Hence, these particular picture stories were replaced by other sets which were found to cause less confusion for the main study.

3.2.2 Spot the difference:

This entails identifying differences between two versions of the same picture. Piloting revealed the task was appropriate in terms of content and clarity. However, the learners' turns tended to be very short; they typed simple sentences during the session and preferred to wait for their partner to initiate and direct the conversation.

e.g., Tamara: can you tell me more about the bedside table? What is it made of?

Kareem: wood

Tamara: ok, anything else?

Kareem: a TV on the left side of the chair

It's on a small table

Since one of the main aims of this study is to capture progress in learners' proficiency in terms of CAF, the instructions were duly modified to ask the learner to write longer sentences and give as much detail as they could when describing an item to their partner. Hence, the following was added to the instructions: *for better interaction, try to provide as much detail as you can when describing an item to your partner.*

3.2.3 Information-gap:

This involves the interlocutors holding different pieces of information and therefore they need to exchange information with each other to fill in the missing gaps. One problem was identified while piloting this task: NS/ST interlocutors did not attempt to correct their partners' erroneous utterances which might help in drawing learners' attention to the accurate use of certain forms, especially those targeted in the current research. Hence, the following was added to the instructions delivered to the NS/ST interlocutors in an attempt to surmount this problem:

> "While working on the tasks, please try to help your partner by drawing his/her attention to the mistakes/errors they make with their English that you notice, especially those related to articles, tense, and use of relative clauses".

3.2.4 Decision-making tasks:

This requires the learners to discuss a problematic issue and together with their partners make an appropriate decision based on the given alternatives. A problem which emerged during the piloting of this task was that one of the pilotees seemed reluctant to pre-plan; she said she is not very acquainted with the skill of notetaking. Hence, due to the importance of planning as a variable in this project to examine its effects on L2 development, the learner was given

instructions/samples (see appendix C) regarding effective notetaking in the following session. These, as reported by the learner, were very informative and helped her to take relevant notes that she could use during the interaction. A copy of these instructions would subsequently be given in the first session of the main study to all the SLs who were supposed to pre-plan prior to their online performance.

To sum up, the piloting phase, which lasted for approximately ten weeks, was beneficial as it helped in identifying problems regarding 1) the choice of the tasks, emphasizing to me that the content of the task should be straightforward so that it does not put an unnecessary cognitive load on learners' processing abilities, 2) lack of clarity in the instructions, revealing that since most of the learners are not familiar with this learning mode, detailed instructions should be given to avoid misperceptions, 3) the appropriacy of the CAF measures; not all the adopted measures were suitable for the context of this study and thus some modifications needed to be made to these measures (all discussed below in **3.3.8.1.1**), so that they became more appropriate for the context of the current study, and 4) the interview questions, highlighting the need to avoid asking leading or biased questions and replace them with more neutral framing as otherwise they could indicate I am expecting a certain answer from the interviewee.

3.3 The Modified Design of the Current Study:

Having carried out the pilot study and addressed the problematic issues regarding the instructions, the tasks' content, etc., I started collecting data for the main study. The following sections will first justify the research design employed for the conduct of this study. Information about the participants/context, experimental design, and materials will be presented afterwards as well as all the modifications which were incorporated. The final section will then discuss the procedures followed for data collection, coding, and analysis.

3.3.1 Justification of the Longitudinal Research Design

Dörnyei (2007) describes longitudinal research as "the ongoing examination of people or phenomena over time" (p. 78). This kind of research is very important, especially when it comes to L2 learning, a dynamic process that typically happens over time. A longitudinal study, therefore, seemed to be ideal given the aims of the current project: to analyse and track SLs' L2 progression due to their online interaction with more proficient L2 interlocutors over a certain period of time. However, it has been argued that the downsides of this approach have prevented many researchers adopting it, instead preferring to investigate L2 performance "at isolated points in time" (Thompson, 2014: 95). These downsides include lack of commitment on the part of the participants due to the long duration of the study as they have the right to withdraw at any time without giving any reasons (Dörnyei, 2007). Different techniques were followed to minimise the occurrence of this problematic issue: firstly, NSs were given a small payment of £100 as compensation for taking part. However, as it was not possible to pay any of the STs or the SLs who were living in Syria at the time of the study, it was hoped that the new experience (i.e., online teaching) and the opportunity to practise their L2 would motivate them to fully commit to this project. Secondly, following Dörnyei's (2007) recommendation with regard to the minimum time required to conduct a longitudinal study within the field of applied linguistics, the duration of each study in the current research was set at six weeks and a month gap was included so that all the participants had a break halfway through the experimental sessions. The one-month gap period was intended to reduce any effects interaction via one mode of SCMC (text chat) may have on the other (voice chat). Finally, understanding that some of the participants might have a tight schedule, they were reassured that the whole interaction would take place online and that the sessions would be arranged at their best convenience, i.e., interlocutors would be free to do the sessions anywhere as long as they had their personal devices with them. It is worth noting, however, that all the interlocutors

did the sessions within the same timeframe: one session of written/spoken online interaction per week.

Another reported disadvantage of this approach is the time and effort required to conduct a longitudinal study, let alone the time required to analyse the resulting amount of data afterwards (Samuda & Bygate, 2008). One way to address this issue was by using computer-aided software (e.g., SPSS, Python), which according to Dörnyei (2007) would free up the researcher's time and avoid data overload. I also tried to collect the data over a certain time period and analyse learners' production immediately afterwards; doing so meant that a large amount of unanalysed data did not build up. The longitudinal nature of the research might also potentially affect the participants' online behaviour (known as panel conditioning effect) due to the experience they would gain from the previous regular online meetings, and hence the participants might react differently in order to please the researcher (Dörnyei, 2007). Therefore, none of the participants was aware of the main aims of the study; they were only informed that I intended to examine online interaction during collaborative tasks. Besides, all the interview questions were neutral (i.e., unbiased) in nature, and different task types/contents were employed so that the learners would be least affected by their performance in the previous session(s).

3.3.2 Justification of Mixed Methods Research

It has been argued that an important decision that a researcher needs to make is related to the kind of research approach they wish to adopt that best addresses the research problem (Senior, 2007). They have to choose whether the research to be conducted is either towards the quantitative or qualitative end of the research continuum. To put it another way, the researcher may intend to investigate hypotheses and present results in numerical terms, or the research project itself may focus more on describing and explaining a particular phenomenon in detail.

Both approaches indeed have their roles and/or advantages in shaping theories, providing distinct perspectives, yet they also have their limitations. Richards (2003) contends that quantitative research, which is based on numerical analysis and statistics, is "not designed to explore the complexities and conundrums of the immensely complicated social world that we inhabit" (p. 8). In other words, quantitative data collected only from questionnaires and tests could provide us with valuable information; however, it would be highly unlikely for quantitative data to be capable of giving details or providing an in-depth understanding to explain the occurrence of any unexpected results. Likewise, when studying a particular context with a limited number of participants via means of qualitative methods, it would be implausible to generalise the findings to a wider context. Shannon-Baker (2015), therefore, maintains that "the implications of using limited approaches in any line of inquiry result in investigating a problem from only a single angle" (p.36). As a result, Creswell & Clark (2017) encouraged L2 researchers to incorporate both qualitative and quantitative methods into longitudinal research as this allows for the limitations of each method to be offset by the strengths of the other in a complementary manner. That is, the researcher should aim to find out how the integration of both methods could provide a more comprehensive interpretation of the researched problem and foster the development of a particular theory. Plano Clark & Ivankova, (2016) also emphasised that mixing two methods can "develop more effective and refined conclusions by using the results from one method (qualitative or quantitative) to inform or shape the use of another method (qualitative or quantitative)" (p. 86). This was noticeable in the piloting phase when interviewing the pilotees about the planning strategies they adopted to prepare for their task performance: the qualitative findings revealed that the social and cultural backgrounds of the learners influenced their planning decisions. That is, those who were concerned about their errors and wanted to produce accurate and complex sentences made greater use of the planning time compared with the pilotees who had a strong orientation towards communication. The

findings, which seemed to be in line with Ortega's (2005), helped explain the pilotees' linguistic performance during the online sessions as well as the subsequent post-tests, provided a complementary perspective to the initial quantitative data, and consequently improved the validity of the research as a whole.

Accordingly, a mixed-methods approach was embraced in the current study, which involves the triangulation of quantitative (i.e., pre-test/immediate and delayed post-tests) and qualitative data (i.e., chat logs, interviews, and weekly reports), so that any bias revealed in one source could be counterset by the other. In other words, using quantitative analysis would provide preliminary results needing further exploration, yet adding a subsequent qualitative element to the study would provide more in-depth information (Creswell & Clark, 2017). As shown in the example above, the data collected from the online sessions/tests alone provided me with a general picture of the learners' progress; however, it did not give me details about what factors affected their planning decisions and L2 performance, or how they improved eventually. That is, were these decisions/improvements influenced by learners getting used to the online atmosphere, the L1 status of their interlocutors, or some other social/cultural factors? Or due to learners attending to the gaps in their interlanguage and trying to make use of their partners' feedback? Neither would the online data tell me the extent to which learners thought they benefited from the online experience. This was further emphasized by Yin (2014) who contends that mixed methods designs "can permit researchers to address more complicated research questions and collect a richer and stronger array of evidence than can be accomplished by any single method alone" (p.66).

3.3.3 Participants & context

Recruiting participants who were willing to fully commit to the study for the duration of five consecutive months, especially native speakers of English, was one of the greatest hurdles that I faced as I started collecting data for the main project. Besides, the burden of working with a large number of participants at the same time was quite onerous, given the amount of work needed and the time to be spent with each interlocutor. Therefore, my supervisor and I agreed to eliminate one of the planning conditions (i.e., the unguided planning group) and decrease the number of the participants to twenty SLs instead of thirty, five NSs, and six English STs. However, we also agreed that the number of participants could not be cut down any further after consultation with statisticians, since doing so would have put the statistical calculations of the results in jeopardy, and that there would then have been insufficient numbers of participants to perform appropriate statistical tests.

Hence, twenty intermediate-level adult SLs were randomly assigned to work with either a NS or a ST on different task types over the duration of twelve weeks of text-based and voice chat, with the intent of inducing SLs to produce more complex, accurate, and fluent structures (see **tables 4 & 5** for key information about the participants).

Table 4: SLs' background information

Groups	Learners' pseudonyms	Gender	Age	Mean age of SL2 per group		Study level
				M	*SD*	
Treatment group	1. Samo	Male	23			Undergraduate
	2. Tala	Female	25			Graduate
	3. Anas	Male	30			Graduate
	4. Ameen	Male	22			Undergraduate
	5. Laith	Male	23	24.4	2.27	Undergraduate
	6. Manar	Female	25			Graduate
	7. Sima	Female	27			Postgraduate
	8. Sana	Female	25			Postgraduate
	9. Majd	Male	22			Undergraduate
	10. Lara	Female	23			Undergraduate
Control Group	1. Abd	Male	20			Undergraduate
	2. Layan	Female	24			Graduate
	3. Jana	Female	20			Undergraduate
	4. Mohanned	Male	25			Graduate
	5. Masa	Female	29	23.8	3.08	Graduate
	6. Natali	Female	23			Undergraduate
	7. Mariam	Female	25			Graduate
	8. Nada	Female	27			Postgraduate
	9. Rima	Female	25			Graduate
	10. Noha	Female	20			undergraduate

Table 5: NSs/STs' background information

Names	L1 status	Gender	Age	Teaching experience
1. Sally	ST	Female	27	Three years
2. Rana	ST	Female	28	Two years
3. Kate	NS	Female	26	Two years
4. Rosy	NS	Female	30	Five years
5. Rami	ST	Male	26	Three years
6. Melanie	NS	Female	26	Three years
7. Anna	NS	Female	30	Four years
8. Fatimah	ST	Female	27	Four years
9. Nadeen	ST	Female	34	Two years
10. Emily	NS	Female	26	Two years
11. Sima	ST	Female	25	Two years

Since one of the aims of the study is to compare how SLs perform under guided pre-planning

conditions, they were divided into two groups of ten dyads. The first group operated under

guided pre-task planning conditions, where the learners were given guided instructions regarding their language use as well as ten minutes prior to the commencement of each online session to take notes related to the assigned task. Then as soon as the ten minutes was over, they started interacting with their partner. The second group, however, acted as a control group, i.e., the SLs were given no guided instructions or time to plan their performance and instructed to start the online session with the other interlocutor immediately. NSs/STs working with the treatment group were asked to provide feedback on learners' errors whenever they deemed this necessary and to encourage the learners to focus on/integrate particular language forms (such as relative clauses) into their output. Neither of these things were mentioned in the instructions given to NSs/STs working with the control group.

The SLs, seven males and thirteen females, ranged in age from twenty to thirty years old. At the time of data collection, they were either still students at a large Syrian university or were graduates. All the SLs were recruited in this study based on their IELTS or TOEFL scores and only those who scored 6 or 6.5 in the IELTS or equivalent in the TOEFL test were asked to take part, so that their levels were all quite similar. During the interviews (conducted either in the piloting phase or in the main study), the learners reported that they rarely had an opportunity to practise English in their contexts, let alone talk to native English speakers. This and the fact that they were able to participate free of charge made some of the SLs very motivated to take part in the study; they mentioned how expensive it is nowadays to enrol in an English course in Syria and saw the benefit of taking part in the study as enabling them to practise their English over an extended period of time. Besides, the learners mentioned that practicing English by going on the Internet every day, watching YouTube, for instance, was not always an option for them given the intermittent Internet connection in Syria. Some learners also mentioned that they had experienced online learning before when they enrolled in English-medium chatrooms, but then quit soon after as they did not find these chatrooms very safe to be in; they were

apprehensive about their privacy. Most of the learners revealed that they have future plans to pursue their studies or find a job abroad, and hence improving their English proficiency is a priority. Since the learners were still living in Syria during the experiment, that is, daily or frequent use of English was not conceivable given their context, it was assumed that any improvements in their L2 proficiency were predominantly due to their online interaction.

Since the tasks required them to comment and give feedback on SLs' production, all the NSs and STs (nine females and one male ranging in age from twenty-five to thirty-four years old) were required to have a minimum of two years' teaching experience to be eligible to participate in the study. It is worth mentioning that the rationale for recruiting both NSs and STs was twofold: firstly, I intended to offer SLs varied opportunities to benefit from the online interaction, interacting with either a NS or ST partner. Besides, I also aimed to investigate whether the L1 status of more proficient L2 interlocutors would have any impact on learners' L2 production and/or anxiety level. Pairing was done based on the timetable of each interlocutor. As far as their computer skills are concerned, all the participants in this study were found to have approximately the same experience of using computers, ranging from one to four years. Almost all of them used Skype and considered it as a convenient tool for chatting.

The consent form[14], approved by the University of Sheffield's School of English ethics reviewers, was sent to all the participants via email/Facebook Messenger. In order not to affect the way they would react in the online sessions, the participants were not aware that I intended to observe their L2 development; they were simply told that the study aimed to examine online interaction during collaborative tasks. All the participants read the consent form and sent a signed copy back. Anonymity and confidentiality issues were taken into account and pseudonyms were used to further mask participants' identities.

---

[14] See Appendix A

3.3.4 Experimental Design

A longitudinal study incorporating a pre-, post-, and delayed test design was carried out over the duration of five consecutive months (**Figure 6** below illustrates each stage of the modified research design).

Figure 6: The modified design of the current study
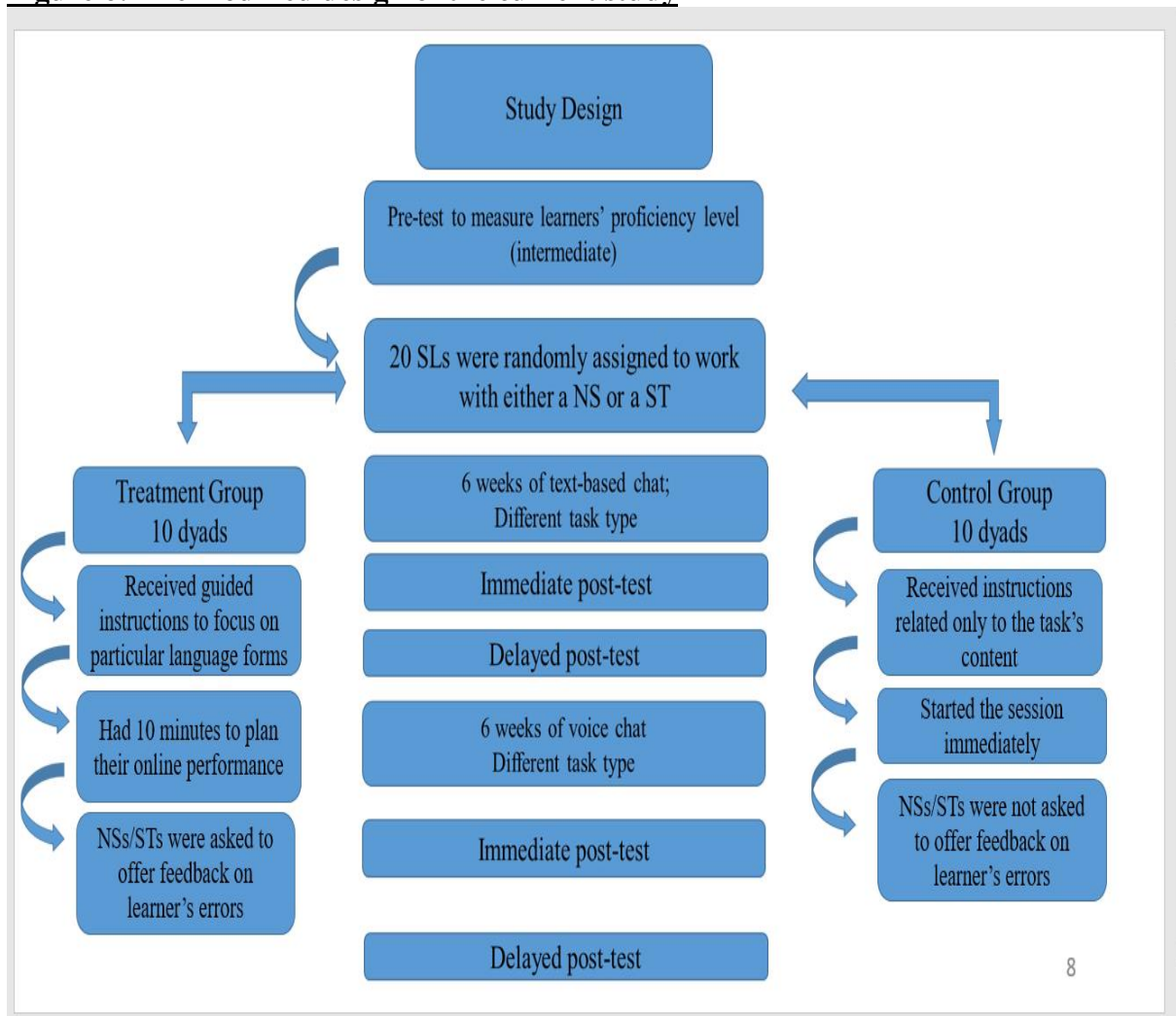
Different cognitively-demanding tasks were implemented during twelve weeks of text-based and voice chatting sessions (see **table 6**). All learners worked with both modes, and all had to begin with the text chat sessions before moving to the voice chat mode; it was believed that starting with the former would be less intimidating for the learners and would give them some time to get to know their L2 partners, whom they had not met before. In evidence of this, when asked what she thought about interacting via text chat, Rama, one of the pilotees, replied:

"I was feeling very comfortable when doing the task; I had enough time to write and edit the sentence before sending it to my partner. I could notice what I did right and what I did wrong as I can see the messages on the screen."

I felt that the real challenge, however, was whether the learners would be able to perform similarly during the voice chat mode which does not have the same affordances as the text-chat environment. Surprisingly, Rama, who frequently expressed her concerns about her speaking skills did very well during the audio chat session, being very confident and relaxed. She mentioned that the text-chat interactions gave her a space to practise her English and build a rapport with someone she did not know, which in turn made it easier for her to participate in the audio sessions held afterwards.

Table 6: Study schedule for all learners

Text Chat	
Pre-test (written & oral narrative)	
Week 1	Spot-the-difference
Week 2	Information-gap
Week 3	Decision-making
Week 4	Narrative
Week 5	Decision-making
Week 6	Decision-making
Post-test 1 (written & oral narrative) + 1 month break	
Delayed test 1 (written & oral narrative)	
Week 1	Spot-the-difference
Week 2	Decision-making
Week 3	Narrative
Week 4	Decision-making
Week 5	Decision-making
Week 6	Decision-making
Post-test 2 (written & oral narrative) + 1 month break	
Delayed test 2 (written & oral narrative)	

A combination of morphological and syntactic measures, which were used in previous CAF research, was adopted in the current study to gauge the quality/quantity of the language produced by SLs during the pre, post- and delayed tests (see **section 3.3.8.1** for a full discussion

of these measures). At the beginning of this study, the learners had to undertake a pre-test, which involved working on two narratives (written & oral). Twenty SLs of an intermediate level were then randomly assigned to work with either a NS or ST under either the guided pre-planning condition or the no-planning condition for six weeks and practised different task types by means of text-based chat (see **table 7).**

Table 7: pairs/groups division

Pairs	Planning condition
1. Samo & Anna	Guided pre-planning
2. Tala & Emily	Guided pre-planning
3. Anas & Sally	Guided pre-planning
4. Ameen & Rami	Guided pre-planning
5. Laith & Rosy	Guided pre-planning
6. Manar & Kate	Guided pre-planning
7. Sima & Sally	Guided pre-planning
8. Sana & Melanie	Guided pre-planning
9. Majd & Rosy	Guided pre-planning
10. Lara & Kate	Guided pre-planning
1. Abd & Rana	No planning
2. Layan & Sima	No planning
3. Jana & Nadeen	No planning
4. Mohanned & Emily	No planning
5. Masa & Fatimah	No planning
6. Natali & Sima	No planning
7. Mariam & Nada	No planning
8. Nada & Rami	No planning
9. Rima & Rana	No planning
10. Noha & Melanie	No planning

Having completed all the text chat online sessions, SLs then carried out an immediate post-test and a delayed post-test (a month later) to detect any short-term/long-term gains (i.e., uptake) of the targeted forms addressed in the pre-test as well as the online sessions. A similar routine was followed a month later with the same groups of learners in the form of audio chat interaction.

3.3.5 Applications used for data collection

a. Skype

All the online interaction between the interlocutors was conducted via Skype. Skype is a multi-party and instantaneous text messaging software program. The accessibility and multimodality of this program made it a perfect tool to achieve the goals of this project. During the text-based chatting session, the conversation window was split into two parts: the bottom part was used to type and send messages, whereas the top part was specified to read received messages (see **Figure 7**). Skype does not only offer a text-messaging feature, but also a range of audio and graphic affordances (e.g., emoticons) that help users express their feelings and perhaps, to some extent, compensate for the lack of facial expressions available to interlocutors in this medium.

Figure 7: A screenshot of learners' interaction via text-based chat
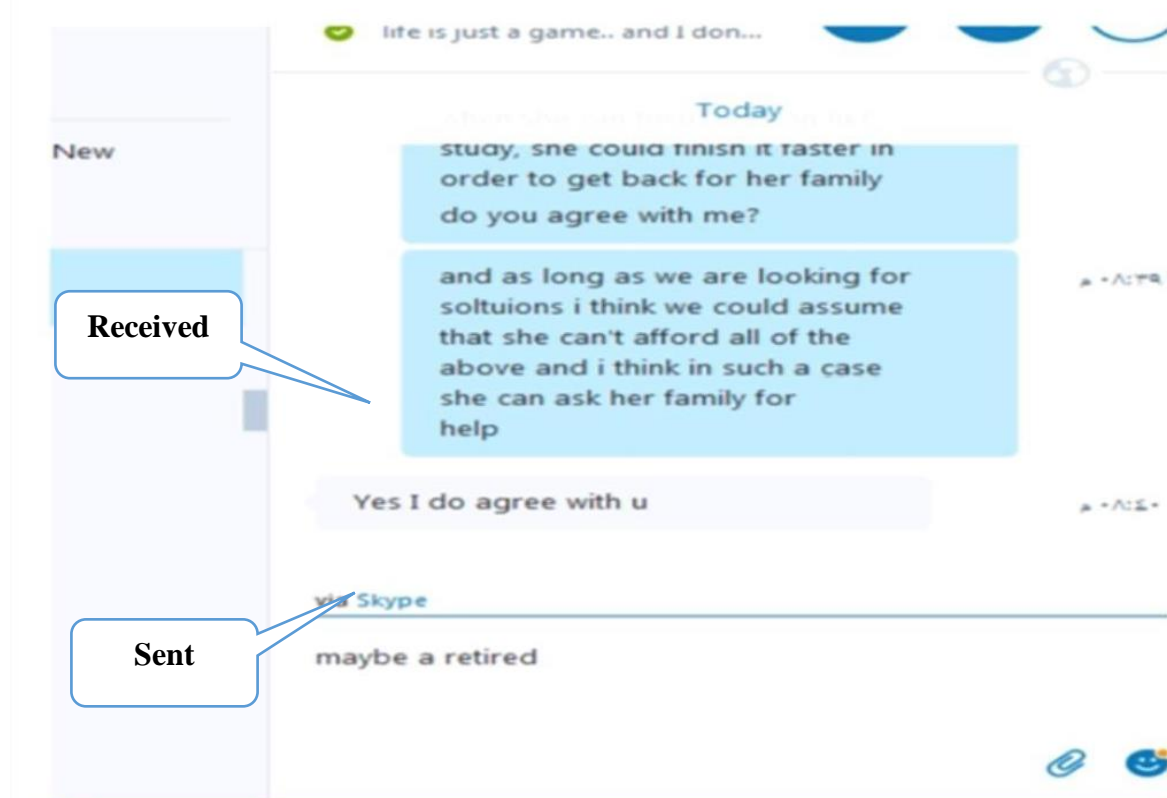
b. Teamviewer & Anydesk

Each text-based Skype session was first videotaped using the Teamviewer software program, which enabled me to share the participants' screens instantaneously. However, as this software

was blocked in Syria after a few sessions of data collection, it was replaced with another program called Anydesk that has similar features to Teamviewer. The SLs were provided with the link to download this program and step-by-step instructions regarding how to set up the software on their computers. At the beginning of each session, SLs were required to provide me with the 'partner ID' or 'address code' that appeared on their screens as soon as they launched the program, so that I could get access to their screens. Though traditional video recordings might provide more information (e.g., learners' facial expressions while working on the tasks), incorporating such methods could have affected learners' interaction because of the intrusive effects of being filmed and subsequently lessened the reliability and validity of the gathered data. Screen capturing programs, on the other hand, allow researchers to collect lots of relevant data unobtrusively (Lai & Zhao, 2006; Lim, 2002; Sauro & Smith, 2010).

It is worth mentioning that getting the learners' consent to use this kind of software (i.e., Anydesk), which gave me full control over their devices, was not a straightforward process. Despite the fact that they were very excited to practise their L2 and improve their English, some of the SLs declined to take part in this study as they mentioned that they had important data on their devices and could not risk sharing their screens with anyone. Alternatively, they offered to videotape the sessions themselves and send me the videos afterwards, but this offer was declined. I felt that there was a risk that, during the session, learners might be too involved in the conversation and forgot to turn on the recording or check if it was working at all. They might also go off-topic and move the conversation in a different direction. For example, during the piloting sessions, the participants were instructed to spend two to three minutes at the beginning of the first online session to introduce themselves to their partners. However, I found out that one of the dyads, who offered to videotape the session themselves, unexpectedly spent about fifteen minutes talking about themselves and not paying attention to the fact that time

was passing and that they still had not started working on the assigned task. Bearing in mind that the sessions could not be repeated twice with the same learners, it was important that I had direct control over each session in order to ensure I could keep each session on track. That is, whenever I was feeling that the participants were spending too long discussing a particular point, I typed them a message to remind them how much time was left.

To address the concerns of the SLs regarding the use of this software (i.e., Anydesk), they had all been assured that they would be free to terminate the connection if they found any suspicious or worrying action on the screen. To put it differently, some of the participants were apprehensive about the fact that as soon as I logged into their devices via Anydesk, I would have full control over their screens and that I could access any of their personal files. Should any of this have happened, however, learners would be able to see what I was doing and would have the right to disconnect their devices. They had also been informed that their devices would only be accessed during the online sessions and that for security reasons they were free to change the ID or the code after the end of each session. However, these precautions failed to convince some of the potential participants to take part in this study.

3.3.6 Tasks

The majority of previous CAF research (e.g., Ahmadian, 2012; Asgarikia, 2014; Mochizuki & Ortega, 2008; Thompson, 2014) investigated L2 proficiency via means of employing mainly monologic narrative tasks; the learners were given a set of pictures and asked to tell a story about these pictures. The current study, on the other hand, examined the effects of longitudinal online interaction on L2 development. This necessitates the execution of tasks that allow communication between the learner and a more proficient L2 speaker; that is to say, both should have an equal amount of information in order to achieve the task goal. Finding and designing

appropriate tasks for the participants to work on during the online sessions, therefore, was a very demanding stage of the research as I had to carefully choose topics and themes that would be of interest to most of the participants, particularly pairs coming from an entirely different cultural background.

Different collaborative learning activities (spot the difference, information-gap, narrative, and decision-making tasks) were used for this project (full versions of all tasks can be found in Appendix B). The rationale behind implementing varied types of activities was based on the claim that tasks vary in their cognitive demands and hence generate differential learning opportunities for focus on form (FoF: narrative and decision-making) and focus on meaning (FoM: spot the difference and information-gap) in F2F and SCMC environments (Swain and Lapkin, 2001). This claim was recently substantiated by the findings of Michel et al. (2019) which demonstrated that "each task type has the potential to trigger the use of specific target structures – but at the same time carries the risk of not eliciting other structures" (e.g., narrative tasks trigger the use of simple past) (p.143). Therefore, they concluded that L2 progress necessitates the integration of different task types in all phases of language development. It is also worth mentioning that given the long duration of the project and the number of sessions the participants had to conduct, it was not possible to ask them to work each time on the same task type as this risked boredom. Doaa, one of the pilotees, commented: "I think the inclusion of different task types have just made the experience like more stimulating and made me look forward to each session. In fact, that was absolutely thrilling".

The selection of the tasks was motivated by previous SCMC and F2F studies (e.g, Yilmaz and Granena, 2010; Hsu, 2012; 2017). Narrative tasks, in particular, have been frequently used in research investigating CAF and been found to work quite well (e.g., Asgarikia, 2014; Mochizuki & Ortega, 2008; Seyyedi et al., 2013; Thompson, 2014). In addition, all the tasks I included in the main study satisfy Ellis's (2003) criteria for a language task that is deemed

conducive to language learning, i.e., a task having a clearly defined outcome, learners focusing on meaning while making use of their own linguistic resources. The task topics were designed to be mostly familiar to the general life experience of the participants. The tasks are discussed in full below.

3.3.6.1 Spot the difference tasks:

Two tasks of this type were used in the current study, one in each mode of interaction. As stated earlier, this is considered as a meaning-focused task that involves giving both the learner and their partner a different version of the same picture and each interlocutor should communicate what s/he can see in their picture, thereby identifying the differences between the two versions. When asking the pilotees to sequence the tasks based on their difficulty (i.e., from simple to complex) at the end of the piloting phase, almost all of them reported that spot the difference tasks were the easiest as all that they had to do was to describe the content of the picture to their partner. Hence, it was presumed that starting the first online session in each mode (text-based chat and audio chat) with a spot-the difference task would be useful for both the learner and the NS/ST for the following reasons: first, the participants might not be very familiar with online teaching/learning and therefore working on a more complex task type could adversely affect their overall performance; second, the participants did not know each other prior to the online sessions and might need some time to build rapport before moving on to discuss more complex tasks in terms of type and topic. One more reason for beginning with this type of task even in the voice chat mode (i.e., after the participants had already got used to the online environment and presumably built a relationship with each other) was the affordances of the mode itself. During the text-based chat, the participants (the SLs in particular) view their partners' messages and have time to type their own. However, in the audio mode, the immediacy of the interaction means SLs have less time to think of what they

would like to say and understand what their partner is saying. Therefore, I believed that it would be preferable to start with less cognitively demanding tasks first.

3.3.6.2 Information-gap tasks

The second task type used in this study involves the participants interacting together in order to fill in the missing information or the gaps and achieve the task's goal. One version of the pictures was given to the SL and the other was given to the NS/ST (see **Figure 8**). The description of this task was provided to the learners as follows:

> *The picture you have is a drawing of Richard's student room at university. Your partner*
> *also has a drawing of the same student room, but his/her drawing is incomplete. Hence,*
> *you need to help him/her complete the drawing by saying where the things go.*

Figure 8: information-gap task (version A & B)
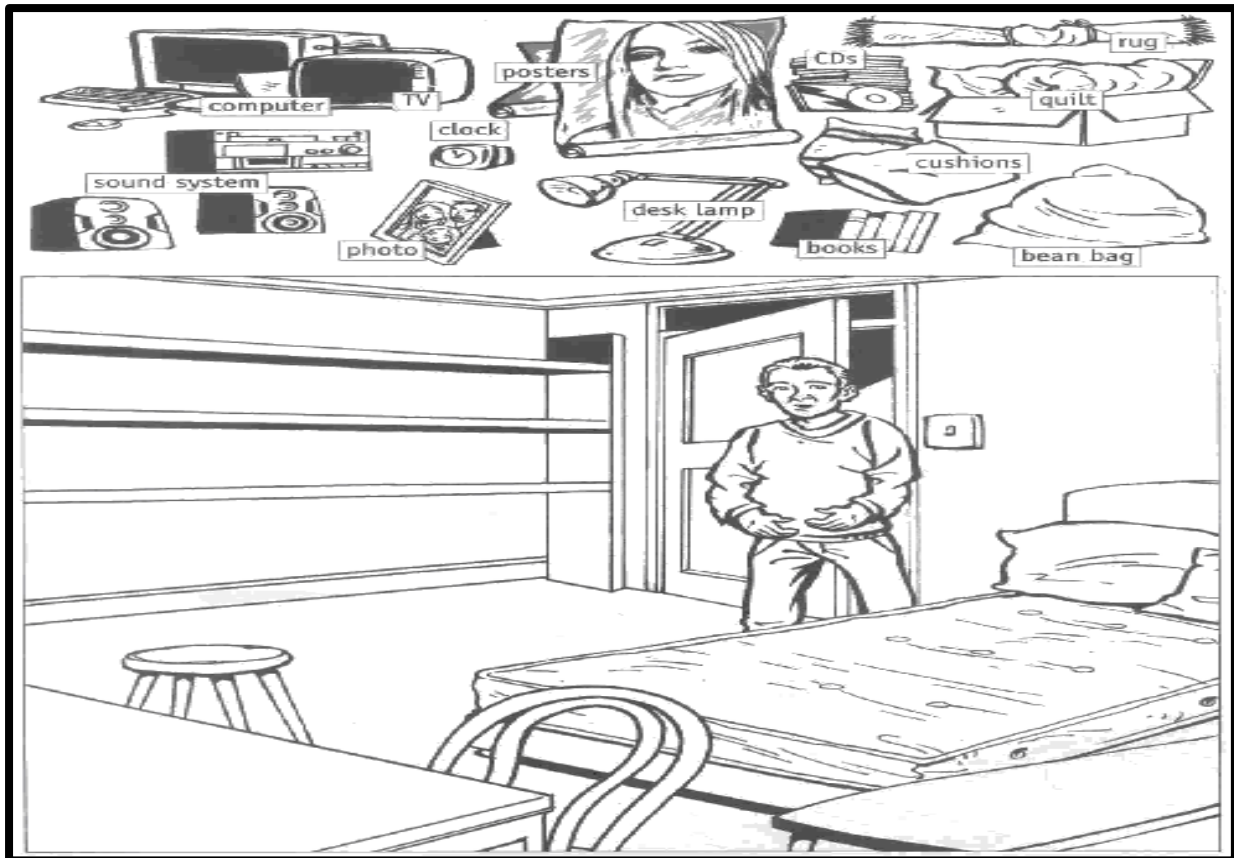
Learners in the piloting sessions had very positive attitudes towards this type of task, describing it as 'useful', 'challenging' and 'interesting'. One of the pilotees commented: "I loved what I was doing, I did enjoy every single part of that task and I wished I had more items and things to describe. Honestly, I never wanted the session to end".

The pilot study also revealed that when the NSs/STs were provided with the version that has all the missing information, the learners had a very passive role and did not get the chance to speak during the session. Therefore, and in an attempt to push the SLs to better communicate with the NS/ST, for the main study learners were given the version which has all the items and were asked to describe these items and their location in the picture to their partners. It is worth mentioning that this task type was only employed during the text-based chat as the participants were still new to online interaction and needed more time to cope with the new learning/teaching environment.

3.3.6.3 Narrative tasks

As stated earlier, this task type was frequently employed in previous CAF and SCMC research (e.g., Ahmadian, 2012; Thompson, 2014), where the learners were given a set of pictures and were asked to produce monologic narratives. However, for the purposes of this study that focused on the effects of online interaction on L2 development, some modifications were made to the design of this type of task to be used during the online sessions. Working in dyads, the participants were given a set of six or eight pictures, three or four each, with which they were required to first take turns to talk about the content of each picture, and then after all the pictures in the possession of both interactants were fully described, they needed to reorder these pictures to make a comprehensible story. Thus, in order to increase the cognitive demands of the task, the pictures were distributed between the interlocutors in a jigsaw and jumbled format (i.e., each had different parts of the story). Piloting revealed that this decision did not only make the task more entertaining and challenging than giving them both the same set in the right order, but also enabled each interlocutor to have his/her own turn when describing their set of pictures. That is to say, the learners got the chance to produce their own elaborated output while still receiving comprehensive language input from their partners. Another interesting finding uncovered when piloting this type of task was that the jigsaw, jumbled format would either force the participants to scroll up the messages, going back to the description of previous pictures provided by their partners during the text-based chat mode, or repeatedly request their partner to re-describe the content of a particular picture so that the dyad could create a satisfactory story during the voice chat mode. Accordingly, the SLs in the main study would have multiple opportunities to review previous messages, talk more, and hopefully notice the different structures used by their more proficient L2 interlocutors. However, to allow comparisons with previous research, the SLs were requested to perform narrative monologic

tasks (written and oral) during the pre-, post-, and delayed tests. The jumbled order of the pictures was also used to increase the cognitive demands of these tasks across the tests.

3.3.6.4 Decision making tasks

For this type of task, the participants were instructed to address the problem at hand by considering all the possible solutions, and then to think of and discuss the pros and cons of each solution, so that they could finally decide on the best alternatives, i.e., the learners needed to prioritise solutions and justify their decisions. All the topics presented in this study were related to everyday situations, addressing smoking, bullying, and workplace dilemmas. The following is an example of one of the decision-making tasks[15] used in this study:

> *Lind has just returned to school and has been out of the study habit for 7 years. She has found it very difficult to get back into the habit of studying. Her time is further stretched by responsibilities of being a wife and a mother of two pre-school aged children. Based on the scenario, together with your partner **identify the problem(s). Discuss all possible solutions** as well as **the pros and cons (if any) for each solution**.*

What distinguishes this type of task is the fact that it promotes learners' communicative and strategic competences when discussing real-life situations. In other words, this kind of open-ended task promotes interaction as learners need to go beyond their L2 repertoires and employ a variety of new forms/structures. This could be likely to happen in particular when the task involves vocabulary and thought-provoking topics which are substantially more challenging than students' actual L2 proficiency level. Therefore, the majority of the decision-making tasks used in this study (three in the text-based chat and four during the audio chat mode) were designed to be of this type. These tasks had been trialled during the pilot and all had been found to generate lively discussion; all the pilotees, without exception, had very positive attitudes

---

[15] This task was adopted from Decision Making Scenarios Worksheet | Decision Making | Applied Psychology (scribd.com).

towards this type of task. Although they mentioned that the concept of the task was more complex than previous tasks as it triggered them to produce complex language to discuss the targeted topic, they said that they still preferred the idea of talking spontaneously to restricting themselves to a particular sentence structure. They also claimed that during this task they had the chance to learn a wider range of real-life expressions that could be useful for everyday use.

3.3.7 Data Sources

Data collected from learners' production during the pre-, post-, and delayed tests constitutes the main source for the quantitative analysis to answer the first and second research questions in this study. On the other hand, the data set used for the qualitative analysis comprises chat logs, recorded audio Skype calls, learners' planning notes, stimulated recall interviews with the SLs and semi-structured interviews with the NSs and STs as well as weekly reflective reports; all of which provided answers to the third, fourth, and fifth research questions. Each of these sources is further discussed in the following paragraphs.

3.3.7.1 Pre-, Post-, and Delayed test

All the learners were required to work on two monologic narrative tasks[16] per test. The rationale behind choosing narrative tasks in particular for the pre-, post-, and delayed tests was to allow comparisons with the findings of previous research. Following Robinson's (2003) Cognition Hypothesis, the tasks used in this study were modified to be different in terms of their cognitive complexity (e.g., wrong sequence order, task content). In other words, the intentional reasoning demands of the tasks were gradually increased in order to make greater linguistic demands of the learners and consequently push more complex output.

---

[16]See Appendix B

In the pre-test, each SL was given a set of pictures and asked to write a comprehensible story about these pictures in a Word document within ten minutes. The poor Internet connection as well as the individual differences between the learners in terms of their typing skills were taken into consideration when deciding on the time needed to address the task goal. This time duration was tested in the piloting phase and deemed appropriate given the learners' proficiency level. The learners were instructed to narrate the story in English using as much detail as they could. As soon as the ten minutes were over, the SLs had to stop writing, close the document, and send it to me straightaway. The test was video recorded using Anydesk and saved for further analysis to evaluate learners' proficiency level in terms of the CAF measures that were employed in this study. Then, the learners were provided with a different set of pictures for which they were given one minute to generate another story, this time carrying out the task orally. The tasks utilised in the tests were different in terms of storyline and delivery method. The rationale behind providing the learners with pictures that contained different storylines was to maintain their interest throughout the study (Samuda and Bygate, 2008). A Skype call was made with each learner so that they were pushed to start the narration immediately after receiving the pictures. This enabled me to audio record and save the story produced by the learners as well. In order to avoid any further delays due to the poor Internet connection (and which could negatively affect their fluency scores), I also requested the learners to use another device to record themselves during the speaking tests while narrating the story on Skype and send me the recordings afterwards. It is worth noting that some of the learners took more than one minute to finish the speaking task; however, only the first minute for each learner was considered for data analysis purposes in an attempt to save time and effort and in the interests of consistency. The same routine was followed for all the other tests: learners worked on the same task type, but different sets of pictures were provided for each test. According to Schmitt (2010) such a design that incorporates pre-, post-, and delayed tests

is beneficial as it helps to "determine the effect of the treatment" (p. 155) and to "demonstrate if long-term retention (i.e., learning) has occurred" (p. 156). Similarly, Rogers & Révész (2019) argue that the pre-test serves to ensure comparability amongst groups prior to the treatment, whereas the post-tests help to verify the effects of the treatment on the measured variable(s).

3.3.7.2 Chat logs:

Logs remain the primary data source for many studies that provide researchers with insights into learner interlanguage development (e.g., Alwi et al, 2012; Eslami & Kung, 2016; Yuksel & Inan, 2014). Evaluating chat logs provides an insight into learners' unfolding composing processes (e.g., deletion, self-correction) while trying to achieve certain communicative goals. Additionally, the permanency of these logs makes them valuable sources for the enhancement of learners' interlanguage; learners can revise the errors they make in their conversation by scrolling the screen's cursor forward/backward, enabling researchers to examine how their language use evolves and improves over time (Blake, 2000). Dyads' interactions during the text-based chat sessions were, therefore, logged and collected for further analysis. However, utilising logs as a main source of data has been criticized as they only offer a narrow picture of learner experience. Smith (2003, 2008) contends that although chat logs are helpful when it comes to interpreting online discourse, they seem to miss an essential component of the data. He believes that "it is precisely these "missing data" that may provide the most insight into the potential roles of monitoring, attention and noticing, and pushed output in interlanguage development within a CMC context" (2003: 5). That is, analysing logs alone will not provide a complete picture of learners' interactions. For instance, it is hard to know whether learners have noticed a gap in their interlanguage or even their partners' feedback unless they explicitly report this. Hence, it was hoped that chat logs gathered concurrently, i.e., while the participants were working on the tasks, would serve as prompts for the recall sessions held afterwards.

3.3.7.3 Recorded audio calls

All the audio chat sessions were carried out via means of Skype audio call, and the participants preferred to use their phones (rather than their laptops) since they believed these would be a more convenient medium for oral interaction. As these sessions commenced, I had to create a group on Skype for each dyad that involved the SL, the NS/ST, and myself.

Figure 9: A screenshot of learners' interaction via audio chat
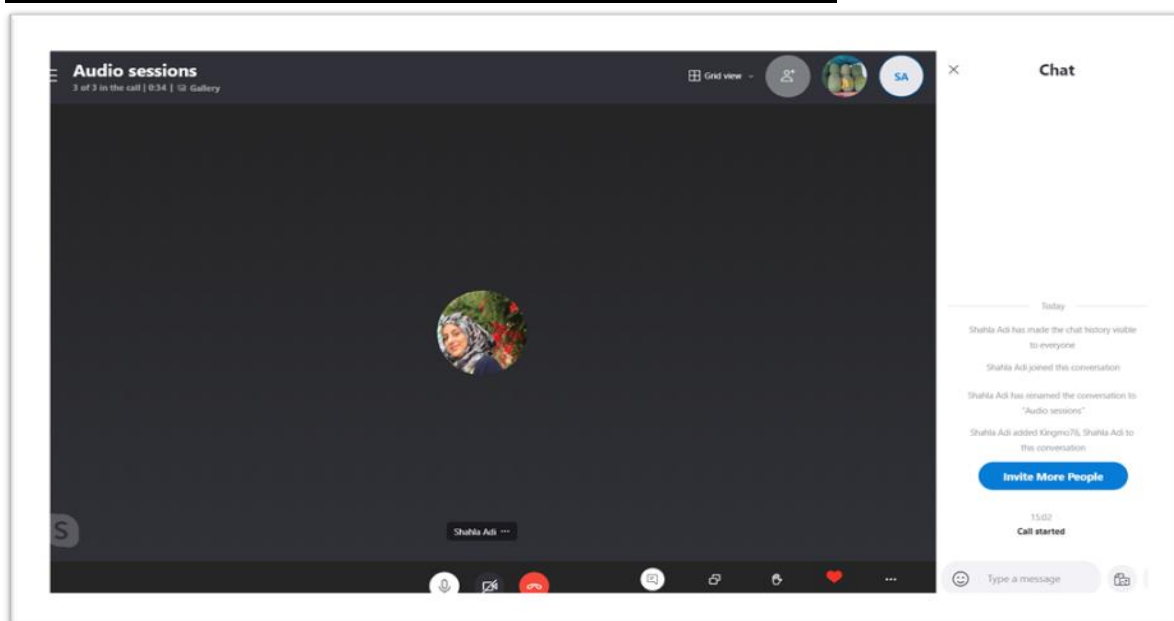
Given the fact that all the data in this study were collected online, it was not possible to place a recorder with every participant to record the session. Additionally, asking the participants to record the sessions using another device or recorder was not an option either as not all the participants reported they had a suitable device for these purposes (i.e., to record a twenty-minute session). I also felt it would have been risky to let the participants download audio recording software onto their devices and record the sessions by themselves. They might forget to turn on the recording or check if it was working at all due to their involvement in the oral interaction with their partners. Therefore, my presence was essential, so that I could record these sessions and save them for further analysis. Although it might sound logical to presume that my presence would have a reactive effect on the participants, resulting in them acting differently (Ericsson, 2002), they had been frequently assured that my role 'is only to record

the session and I will not be interfering in the conversation at all'. The hope was that participants would be so immersed in achieving the task's goal and maintaining the conversation that they would not have time to worry about my presence. It was also important that all participants experienced the same treatment conditions, in terms of the effect caused by my presence during the sessions.

3.3.7.4 Learners' planning notes:

Learners working under the guided planning group were given ten minutes prior to the beginning of each online session to take notes in a separate Word document. The length of the planning time was adopted from the majority of previous CAF studies (e.g., Asgarikia, 2014; Bui & Skehan, 2018; Ellis, 2009; Seyyedi et al., 2013), and was tested in the piloting phase to ensure that SLs had enough but not excessive time to prepare for their online performance. This time duration was, therefore, considered as the threshold for planning during the main study as well.

The importance of learners' notes as demonstrated by the pilot study was that they would provide a clearer idea of the techniques learners follow or the aspects of language (spelling, grammar, content) they choose to focus on during the planning time; and they would be a valuable source to determine whether or not the learners had made use of these notes during the online sessions, hence helping detect the effects, if any, the planning time had on the learners' L2 performance. Following Ortega (1999) and Thompson (2014), the learners' screens were video recorded as they were taking notes which would serve as a visual reminder or as a stimulus for them to retrieve thoughts and comment on their planning strategies during the stimulated recall interviews.

3.3.7.5 Stimulated recall interviews:

Stimulated recall interviews, through which the participants were asked to verbalise their thoughts and provide explanations for their online behaviour following each chatting session, provide an additional data source from the learners' point of view about various aspects of their learning and experiences (e.g., uptake of particular forms). The interview questions[17] were partly adapted from Ortega (1999), Yuan & Ellis (2003), and Thompson (2014), who also investigated learners' L2 production and planning strategies. The questions were aimed to prompt learners to think about what aspect(s) of their language they were focusing on either during the planning time or throughout the online sessions (e.g., form, meaning, word choice). The learners were also frequently asked about their feelings before/during the online interaction as well as their attitudes towards the task they worked on, the mode of interaction, and their partner. The same questions were asked following each chatting session to compare learners' responses with those from the previous weeks.

The recall interviews were also piloted to avoid inappropriate and/or biased questions that might cause reactivity and affect participants' reflective accounts (Dörnyei, 2007; Ericsson, 2002; Smagorinsky, 1989). For example, a question like "How did chatting with a person you've never met make you feel?" was identified by my supervisor as a leading question and hence a more neutral framing was recommended: "Please tell me how you felt interacting with the person you were chatting with". More follow-up questions were then asked to obtain further details. All the interviews conducted with the pilotees were very enlightening, providing interesting information about SLs' strategies to approach the tasks, changes in their anxiety levels, and the techniques their teacher interlocutors used to draw their partner's attention to

---

[17] See Appendix D

grammatical features (e.g., using capitals). The final version of the scheduled interview questions was as follows[18]:

➢ How did you find the instructions/task?

➢ How did you find the planning time?

- Very useful

- Useful

- Somewhat useful

- Not useful at all

- Not sure

➢ What aspect(s) of your language did you focus on when you prepared for the task/during the online session?

- Grammar

- Content/ideas

- Word choice

- Spelling

- Pronunciation

- Something else

➢ Please tell me how you felt interacting with the person you were chatting with.

➢ How well did you feel you worked with your partner?

➢ Could you please tell me how your partner provided feedback to you?

---

[18] It is worth noting that these general questions were asked in addition to other specific ones according to each participant's interaction. For instance, when a specific incidence of error correction was identified, the participant was asked the following questions: "When you were doing the task, I noticed your partner said (an example of correction). Could you tell me about that? What were you thinking about at this time? What was going in your mind when you typed "X"?" Then I asked follow up questions depending on what s/he said, such as how helpful/unhelpful s/he felt it was, whether s/he understood or not, etc.

➢ How helpful did you find your partner's corrections of your English?

➢ Please tell me how you felt interacting via text/voice chat.

Stimulated recalls are considered the least reactive among introspective techniques and therefore are favoured over concurrent introspection such as providing think-aloud commentary, since "the targeted thought processes are not affected by the procedure in any way" (Ericsson, 2002; cited in Dörnyei, 2007: 149). Likewise, implementing stimulated recall interviews has the advantage of being able to elicit more information from the informants than a non-stimulated equivalent format (Dörnyei, 2007), and hence researchers may obtain information they perhaps might not have been able to gather otherwise (e.g., participants' interpretation of events, and their thinking at a particular point in time). Therefore and unlike previous (mostly quantitative) research on SCMC, retrospective interviews were conducted with SLs after each chatting session. Despite the fact that conducting the stimulated recall interviews immediately after the chatting sessions could have generated more accurate answers from the respondents as they would have been describing very recent events (Nisbett and Wilson, 1977), one could assume that such a procedure might put an extra load on the participants as they may feel bored or tired after their chatting sessions and consequently perform poorly in the recall sessions. Hence, the recall interviews were held a couple of hours after the chatting session in order to avoid fatiguing the interviewees but not so long after the event that informants would be unable to recall in-depth what was going on during the interactions. This gave me a chance as a researcher to re-examine the data, identify instances where the learners paused, used complex structures, edited their sentences, and to locate some other interesting segments, such as when learner interactants sought their partners' help when they were experiencing difficulties finding the appropriate vocabulary or the right way to describe something. According to Ericsson (2002), "such thorough engagement with the text

can reveal certain subtle issues that require clarification" (cited in Dörnyei, 2007: 149). Aware of the potential problems as well as the limitations of stimulated recall interviews, that is, the delay between the process and report and consequent forgetting, recalls for the chatting sessions took the form of the SLs watching the live recordings of the whole conversation. It is worth mentioning that the interviews were carried out in Arabic with the Syrian informants, so that I could lessen the cognitive load they might encounter while recalling their thoughts.

During the interview, which lasted for approximately thirty minutes and was audio recorded, the SLs were invited to share my laptop screen, using Anydesk, so that they could review their chat logs. They were prompted to talk about interesting segments of their online interaction (e.g., pausing, self-correction/editing) and planning strategies as they were reviewing the chat logs. I paused the recording at these instances and invited the participant to recall what s/he was thinking. Moreover, in order not to miss other instances that might be of interest to the participants, they were invited to stop the video whenever and wherever they found something that they would like to comment on (e.g., strategies used by NSs/STs to examine whether the SLs understood the feedback correctly). It was hoped that video recordings would give participants a very strong stimulus to recall their relevant thoughts and consequently enable me to obtain more in-depth data concerning what was going on during the interaction (Gass & Mackey, 2017; Mackey & Gass, 2015). Smagorinsky (2008) argues that triangulating concurrent and retrospective accounts helps in capturing "the cognitive and, inevitably, social processes involved in the participants' analytic and interpretive work" (p.396). Stimulated recall interviews might also facilitate learners' uptake of the feedback they received from their partners while the online interaction was in progress; they make the noticing of these sequences more effective in terms of accuracy since participants were supposed to explain how they

perceived their interlocutors' feedback and doing so may have enhanced their metalinguistic awareness.

In contrast to the learners' stimulated recall interviews, semi-structured interviews[19] were carried out with all the NSs and STs at the end of this study to further explore their perceptions of the online experience. They were also asked about the implemented tasks/instructions and the feedback approach they adopted (i.e., their choice to provide explicit/implicit feedback, or maybe a mix of both). The main questions asked to the NS/STs during the interview were:

➢ Please tell me how you find the online experience.

➢ Please tell me how you felt interacting with your partner via text-based/voice chat.

➢ Please tell me how you felt interacting with the person you were chatting with.

➢ How well did you feel you worked with your partner?

➢ Please tell me how you found the tasks. Why?

➢ How effective do you think the feedback you gave was? Why?

➢ Why did you try to avoid explicit feedback? Could you tell me more about this, please?

➢ Tell me about the moments during the online exchanges you felt were particularly helpful or confused.

It was hoped that the participants' responses would provide useful insights for researchers, teachers, and those responsible for curriculum development.

3.3.7.6 Reflective reports:

SLs were also required to write a weekly report in Arabic to evaluate and reflect on their performance. Although none of the previous SCMC studies have asked learners to write

---

[19] See Appendix D

reports, it was hoped that such reports would offer a quiet space for SLs to think reflectively without being overwhelmed with chat logs or interview questions. Reports written by the pilotees in the pilot study uncovered important data regarding learners' awareness of the progress they were making in their production (i.e., evidence of relatively long-term uptake of some forms). Besides, the reports had the potential to reveal further information that pertains to changes in learners' anxiety levels: their feelings before, during, and after the session. This procedure was therefore copied in the main study.

To sum up, while the findings of the pilot study led to a number of amendments related to the design of the main project, the challenges faced while collecting the data (e.g., SLs dropping out of the study, time spent on data collection) resulted in recruiting fewer people than originally planned as well as excluding one of the pre-planning conditions. Additionally, due to the fact that some participants dropped out of the study in the middle of the experiment, it was not possible to gather all the data at the same time. Therefore, data collection procedures were split into two parts, i.e., working with one group of participants at a time. I also had to frequently reschedule the online sessions because of the intermittent Internet connection in Syria. Accordingly, data collection procedures took over a year; much longer than expected. It is worth noting, however, that each participant still followed the same schedule of six weeks of treatment, post-test, 1 month break and delayed post-test. The following sections will discuss data analysis and data coding procedures.

3.3.8 Data Analysis

One of the primary aims of this project was to analyse learners' L2 development and to investigate the combined effects, if any, of the mode of interaction and the complexity of the task on their overall performance during the online sessions. Unsurprisingly, the longitudinal

design, the number of the participants as well as the number of variables addressed in the study all resulted in a massive amount of data that needed to be quantitatively and qualitatively analysed (e.g., online sessions, tests, interviews, and reports). Additionally, the intermittent nature of Internet coverage in Syria and the lack of access to a keystroke logger did not allow me to monitor and record each keystroke typed by the SLs; and therefore, it was not possible to calculate fluency measures for their written production. This meant that in order to fully address the above-mentioned aim of the study, I would have needed to manually analyse the text-based online sessions and the written tests of all the SLs to capture any changes in their fluency scores. Due to the extensive amount of data and to time constraints, my supervisor and I did not consider this as a plausible path to take. Therefore, the research questions were modified so that quantitative analysis would only encompass learners' production during the pre-, post, and delayed tests (see **section 3.1**). Thus, this modified research design and focus would still explore the combined effects of the guided instructions, planning time, as well as the effect of the feedback provided to the treatment group during the online sessions on learners' subsequent production throughout the tests. We also agreed that data from a small sample of the learners (i.e., four case studies) would be qualitatively analysed; these include chat logs, stimulated recall interviews, and written reports.

This study aims to analyse how and whether SLs' L2 production improves as a result of their longitudinal online interaction with more proficient L2 interlocutors. It also examines what learners did during the online sessions and what effects, if any, planning had on their performance. Therefore, the integration of both quantitative and qualitative data analysis enriched the analysis and findings of the current study. Computer-aided quantitative software (Praat and SPSS) was used for data analysis purposes, which helped me detect the number and length of pauses made by the learner per test and facilitated the statistical analysis (performed

by means of independent samples t-tests and paired samples t-tests) of the quantitative data. According to Dörnyei (2007) using computer-aided software results "in considerable gains in efficiency" and "frees up the researcher time and helps to avoid data overload" (p. 265). The following sections will further elaborate on the tools used for data analysis.

3.3.8.1 Quantitative Data Analysis

Since research question one aims to examine the combined effects of manipulating task complexity along with pre-planning/no planning conditions during the online sessions on learners' subsequent L2 production, learners' scores in the pre-, post-, and delayed tests (in terms of specific CAF measures- see below for a full discussion of these measures) were compared per group using quantitative methods. Quantitative analysis was also employed to answer the second research question which aims to capture the impacts of the longitudinal treatment received by the learners and detect short-term/long-term CAF gains, if any, in their oral and written production across all the tests. Language development was then determined by the extent of the pre-, post-, and delayed test gains. Independent samples t-test and paired samples t-tests were performed to establish learners' progress.

Quantitative data gathered in the present study will be further discussed in the following sections. I will first introduce the CAF measures that were initially used to analyse learners' L2 production as well as all the changes that were made due to the modified design and research questions. Then I will move to discuss the qualitative phase that was employed to 1) elaborate on the quantitative results and 2) introduce a clearer picture of learners' behaviour as well as learning strategies.

3.3.8.1.1 CAF Measures:

An important question promulgated by previous research is whether CAF components operate in complete independence from each other. That is, researchers question whether the change or rather the improvement in one variable happens at the expense of other aspects of production due to the fact that learners' attentional resources are limited (e.g., Ellis, 1994; Larsen-Freeman; 2006; Robinson, 2001; Skehan & Foster, 1999; Yuan & Ellis, 2003). Ellis (1994), for instance, claims that attending to input while simultaneously trying to monitor output might negatively affect the flow of communication and interfere with fluent production of the target language. In stark contrast to Ellis, Robinson (2001) proposes a different view and considers CAF as non-competitive pools; therefore, manipulating the cognitive demands of a task can help learners instantaneously access multiple features of production. More recently, accumulative evidence indicated that the three components as well as their subcomponents are interconnected, yet they do not necessarily improve collinearly in SLA (Housen et al., 2012; Lambert & Kormos, 2014; Michel, 2017).

Another thorny issue, widely discussed in previous SLA studies, has been whether general (e.g., error-free clauses) or specific (e.g., articles) measures are more valid and efficient as indexes of L2 performance, L2 development, and L2 proficiency (e.g., Crookes 1989; Skehan, 2003). That is, there is a lack of consensus in terms of how CAF have been operationalised. More recent research, however, has called for the incorporation of both general and specific measures (Ahmadian, 2012; Larsen-Freeman, 2006; Norris & Ortega, 2009; Révész, 2011; Zalbidea, 2017). Larsen-Freeman (2006), for instance, proposes that when measuring CAF, L2 researchers should integrate the use of general and specific measures to carefully observe more detailed features of production. Similarly, Révész (2011) also believes that relying only on

global measures "might result in an incomplete description of task effects" (p. 176). Thus, general and specific measures were applied to measure CAF in the present study.

Additionally, Ahmadian (2012) echoes Norris & Ortega (2009) and argues that "using multiple measures for assessing each dimension of performance (CAF) may yield a more valid and comprehensive picture of a construct if and only if the measures used, tap different facets of the construct in question" (p.140). Hence, different measures were employed to assess the linguistic complexity, accuracy, and fluency of learners' L2 production in the current study. It is worth noting that all of these measures were adopted from previous research and were chosen specifically to fit the aims of the current study (e.g., Hsu, 2017; Thompson, 2014; Ziegler, 2018). According to Handley (2014), duplications of the same outcome measures, generally known as 'instrumental replications', are essential to allow comparisons with previous research, enhance the validity of the measures, and consequently "permit the demonstration of the generality of findings" (p. 51).

In what follows, a concise description of the measures used in previous SLA and CMC research will be provided for each component of the CAF triad, which in turn guided the decisions made in the present research to gauge language development.

3.3.8.1.1.1 Complexity Measures:

Previous studies have employed various measures to operationalize linguistic complexity, each addressing a distinguishing feature of L2 production, including syntactic complexity, syntactic variety, and lexical variety/density. Bulté & Housen (2012) argue that these "are hybrid measures which simultaneously tap into several sub-components and subdomains of L2 complexity" (p.29). That is, any progress in this component has been linked with learning new syntactic structures, rules, and vocabulary.

i. Syntactic complexity measure: L2 targeted forms

Syntactic complexity, frequently viewed as the most intensively operationalised constituent of linguistic complexity in SLA research, has been scrutinised as a dependent variable to gauge impacts of manipulations (e.g., task type, planning time) on written and oral production (Kuiken, Vedder, Housen & De Clercq, 2019). A variety of indices have been utilised to assess this sub-component: syntactic complexity has been primarily measured in terms of the ratio of clauses per a general unit, such as AS-unit and/or the mean length of this unit. An AS-unit is defined as a "single speaker's utterance consisting of an independent clause or sub-clausal unit, together with any subordinate clause(s) associated with either" (Foster et al., 2000: 365). Independent sub-clausal units could be a phrase or a longer utterance that can be understood when recovering the ellipted elements from the discourse, as in the example below:

E.g., Alma: How many mugs you have in your picture beneath the cupboard?

   Kareem: Three.

Norris & Ortega (2009) claim that an AS-unit could ensure an acceptable level of reliability and validity compared to other measures, given that they allow the analysis of oral data and inclusion of sub-clausal units. Additionally, an AS-unit is considered "essentially a syntactic one, and syntactic units are genuine units of planning" (Ahmadian, 2012: 140). Furthermore, this unit has been used in previous SCMC studies (e.g., Adams et al., 2015; Hsu, 2017; Sauro, 2012; Ziegler, 2018) to evaluate the syntactic complexity of learners' L2 production in voice and text chat, as the latter also contains traits of oral language. All of the above justifies the use of AS-units in the present study.

Relying primarily on global measures (e.g., mean length of AS-unit) to gauge syntactic complexity (also accuracy and fluency) has, however, increasingly been criticised. Despite the fact that these measures may indicate higher levels of performance, they are not likely to unveil development in syntax or sentence structure at different levels of proficiency, and hence, there has been a call to employ more 'fine-grained' measures (Bulte & Housen, 2012; Housen et al., 2012; Kuiken et al., 2019; Lambert and Kormos, 2014; Michel, 2017; Norris & Ortega, 2009). As previously mentioned, Bulte & Housen (2012) propose a versatile definition of complexity, suggesting that: "A language feature or system of features is seen as complex if it is somehow costly or taxing for language users and learners, particularly in terms of the mental effort or resources that they have to invest in processing or internalizing the feature(s)" (p. 23). An example of these complex features was provided by Bulte & Housen (2012) who contend that English relative clauses are more difficult to produce and hence they are more likely to be acquired later compared with other linguistic forms (e.g., coordinate structures). Accordingly, syntactic complexity in the current study was operationalized via means of examining learners' use of linguistic forms known for either being difficult to produce or as non-salient in oral speech, especially for EFL learners (i.e., relative clauses, articles, third person -s singular). It was assumed that the SLs would struggle to incorporate these forms into their own production as some of them are not part of their L1, or simply because the forms need considerable mental effort to be internalized. Thus, 'the number of relative clauses per AS-unit' was used to measure the syntactic complexity of learners' production and was trialled in the pilot study.

One of the problems which emerged during the piloting was when the guided pre-planning learners were instructed to focus on grammar: "While planning, think about grammar. For example: he likes the pair of shoes which looks trendy". However, learners frequently avoided employing the forms referred to in the instructions (e.g., relative clauses). When asked, one of

the pilotee learners replied: "Just thought you were talking about grammar in general and using expressions like "she thinks", etc.". The instructions were duly modified to: "While planning, think about grammar; in particular, employing forms like "he like**s the pair of shoes <u>which looks trendy</u>**". The forms were thus underlined, accentuated, and made more prominent. However, similar to the findings of the pilot study, most of the learners in the guided planning group avoided incorporating relative clauses into their production even after being instructed and directed to do so during the main study. Possibly they should have received further training on how to employ and integrate different forms of complex linguistic structures prior to the commencement of the study. Therefore, learners' use of these forms during the sessions was only viewed as an indication that some progress was taking place, and that the learners were attending to these forms and attempting to produce more complex language. Other syntactic measures used in this study were related to the average length of AS-units as well as the ratio of clauses per AS-unit (Yuan & Ellis, 2003). Although these measures do not relate as accurately to the previously mentioned definition of complexity in that they are not considered costly or taxing for language users, they might still suggest that the learners are developing and producing more elaborated utterances (Adams et al., 2015; Ahmadian et al., 2015; Hsu, 2017; Sauro, 2012). The following is an example from the pilot study of an AS-unit that contains three clauses, one of which contains a relative clause:

E.g., To the left of the computer, there is a TV which is turned on.

ii.   Syntactic variety

Following Yuan & Ellis (2003) and Ahmadian et.al. (2015), this variable was measured via calculating the total number of different grammatical verb forms the learners produced per task, in terms of tense (e.g., simple past/present), modality (e.g., could, might), and voice (e.g.,

passive voice). It was assumed that the more varied learners' grammatical verb form, the more complex their production.

iii. Lexical variety

Lexical variety was first operationalized via means of the well-known type-token ratio (TTR) (Richards & Malvern, 2002), calculated by dividing the overall number of types (i.e., different words) that occur in a text by the total number of tokens (i.e., running words). However, one problem about using TTR, as identified by previous research, was that the tool is sensitive to the length of the text analysed (Hsu, 2012), which in turn will affect the reliability of the data and generate misleading results. Ortega (1999), for example, maintained that "there is a negative, but nonlinear, relationship between sample size (i.e., number of tokens) and type-token ratio" (p.133). That is, a minimum number of tokens and text samples of an analogous length are essential for the reliability of this measure. Since these two conditions were not met, as per the findings of the pilot study, and narratives of different length were produced by the learners, the measure of textual lexical diversity (MTLD) (McCarthy, 2005) was chosen instead for measuring the lexical richness of learners' utterances. Koizumi (2012) found that amongst other lexical variety measures, including TTR, MTLD was least affected by text length.

iv. Lexical appropriateness

While analysing learners' production across the pre-, post-, and delayed tests in the pilot study, it was found that the pilotees were trying to make use of the new knowledge (e.g., vocabulary) they gained during the online sessions, and use structures/vocabulary they had not used before (probably to show their L2 competencies). However, their attempts were not successful all the time, i.e., sometimes they employed words/structures that were grammatically correct yet inappropriate in terms of meaning, given the context of the task.

E.g., The kids are preparing for a picnic; Daniel is putting the food in the **<u>baggage</u>** (*instead of* **basket**)

It was, therefore, necessary to add another complexity measure to assess the accuracy and appropriacy of the lexical items produced by the SLs. Hence, all the words/structures learners produced throughout the tests in the actual study and which did not reflect the context of the narratives were counted as inappropriate.

3.3.8.1.1.2 Accuracy measures

Accuracy in the current study was defined as "the extent to which an L2 learner's performance (and the L2 system that underlies this performance) deviates from a norm (i.e., usually the native speaker)" and hence "such deviations from the norm are traditionally labelled (errors)" (Housen et al. 2012, p. 4). Following the recommendations of previous CAF research, global and specific accuracy measures were utilised in the pilot study to allow comparisons over different languages, generalise the findings to other contexts, and perhaps identify small changes in accuracy (Housen & Kuiken, 2009; Lambert & Kormos, 2014; Michel, 2017). Accordingly, accuracy measures used in this study comprised percentage of *errors per test* and *percentage of error-free relative clauses per relative clause*; both of which were used in previous studies (e.g., Adams et al., 2015; Thompson, 2014). A typical grammatical error made by the learners in this study would be for example, '*there is a table which <u>have</u> two candles on it*' instead of '*there is a table which <u>has</u> two candles on it'.* Hence, grammatical accuracy in the current study was measured by calculating the number of relative clauses that had no grammatical errors and dividing these against the overall number of relative clauses the learners produced per each session.

Housen et al. (2012) also contend that accuracy is related to "*appropriateness* and *acceptability*" (p. 4). Thus, not only grammatical errors but also communicatively inadequate

use of the targeted forms (i.e., relative clauses, articles, third person -s singular) was considered as deviation from target-like performance in this project. To put it another way, if a learner produced a relative clause, a tense, or any other language form which was grammatically correct but did not reflect the context of the task, it was classified as inaccurate. For instance, while working on the information-gap task, one of the learners wrote:

*The **desk** which has a **poster** on it…* (Instead of) *the **table** which has a **framed picture** on it…*

This sentence was, therefore, classed as inaccurate. Repeated relative clauses were also excluded from the analysis to avoid overuse of the form. This could be seen in the following sentence:

The girl who is wearing, erm who is wearing a skirt and a t-shirt is standing in front of the mirror.

Given that the learner repeated the same structure twice in this sentence, only one instance of using relative clauses was logged when analysing the data. This kind of repetition was mostly associated with the oral mode; perhaps this was a strategy learners followed to give themselves some time to think of what to say next.

Owing to their performance in the piloting phase, more detailed instructions were provided to the SLs to push them to integrate more complex structures (see **section i**). Nevertheless, the learners frequently tried to avoid producing relative clauses as the data from the actual study revealed. Probably they preferred not to use forms that they were not very sure about; they were trying to stay on the safe side and produce sentences that contained as few errors as possible. Another reason could be that when the more proficient L2 interlocutors were providing implicit feedback, i.e., indirectly attempting to draw the learners' attention to the possibility of integrating these forms into their L2 production, the latter were not perceiving this implicit message as such. Rather, the learners might have understood that their

interlocutors were repeating what they said using different structures. Therefore, '*the percentage of error-free relative clauses per relative clause*' measure was modified to '*the percentage of error-free clauses as compared to the total number of clauses produced by the learner per session*' (Ahmadian et al., 2015; Yuan & Ellis, 2003). Following Seyyedi et al. (2013), all morphological, syntactical, as well as lexical choice errors were counted).

3.3.8.1.1.3 Fluency measures

Following Tavakoli & Skehan's (2005) definition of fluency (see **section 2.2.3**), SLs' production in this study was operationalized via different fluency measures, including mean duration of silent pauses, mean duration of filled pauses, speech rate, and mean duration of repairs. As mentioned in **section 3.3.8**, there were difficulties in measuring the fluency of learners' production during the text-based chat sessions as well as the writing tests due to the poor Internet connection in Syria and the lack of access to keystroke logging software. Additionally, calculating fluency measures for learners' L2 production was a challenging endeavour, as unlike in previous research (where the participants were conducting mainly oral monologic narrative tasks), the learners in this study were working with more proficient L2 speakers on different types of dialogic tasks (e.g., Ahmadian et al., 2015; Thompson, 2014; Mochizuki & Ortega, 2008; Tavokoli & Skehan, 2005). This was another reason for choosing to only analyse learners' production across the pre-, post-, and delayed tests, so that I could compare my results with those of previous research.

The mean duration of silent pauses refers to the instances where there was a total silence on the part of the learner (Skehan & Foster, 1999). Hence, the number of silent pauses the learners produced per session was divided by the total speaking time. The second fluency measure gauged the mean duration of filled pauses; instances where the learners used filler words (like

eh, erm) (Mehnert, 1998). Thus, the number of filled pauses the learners produced per session was divided by the total speaking time. Such breakdown in fluency is believed to "reflect the planning and conceptualization stages of language production" (Michel, 2017: 56). It is worth mentioning that in previous research there was a disagreement regarding the pause duration with which we could determine the fluency or otherwise of the learner's L2, with proposals as low as .25 of a second (de Jong et al., 2012; Kormos & Denes, 2004; Révész et al., 2016), while Tavokoli & Skehan (2005) suggest that total or filled silence greater than .4 a second in a face-to-face conversation is conceived as a disfluency in the learners' production. Thompson (2014), however, believes that pausing for more than 1 second, especially when working on a narrative task, is not necessarily an indication of disfluency. Jepson (2005), on the other hand, considered a pause of more than zero to six seconds between turns in voice chatrooms as an indication of lack of fluency due to the distinct nature of online interaction and the lapses between the real-time utterances of the interactants. Jepson (2005) contends though that further research should be done in online environments to further analyse these pauses and determine whether they are related to the technology itself, the learners' L2 proficiency, or some other factor. Accordingly, and since the interaction between the participants in this study was carried out online, pauses greater than 6 seconds were counted as an indication of disfluency in the pilot study. Yet, due to the modifications made to the design of the actual study which would analyse learners' oral performance during the speaking tests only, the pause duration was set back to 0.25 seconds.

Speech rate or the number of syllables produced per minute was another fluency measure that was employed in this study following previous SLA research (e.g., Gilabert, 2007b; Kawauchi, 2005; Sanguran, 2005: Seyyedi et al., 2013; Yuan & Ellis, 2003). According to Michel (2017), the speed of learners' production "is associated with control of and access to proceduralized knowledge" (p.56). Yet, according to Thompson (2014), this measure has been found to be

problematic; syllables could include L1 use or perhaps learners could frequently use the same words/structures and thus appear to be fluent given the overall number of syllables they produced. Therefore, following Thompson (2014), who examined the effects of guided planning and task complexity on oral development, the total number of syllables produced per minute measures was replaced by 'the total number of syllables produced per minute of pruned speech'. That is, speech excluding repetitions, self-corrections, false-starts, L1 use and incomprehensible language.

The final fluency measure was related to the mean duration of repairs (e.g., repetitions, false-starts) with the aim of detecting any changes in fluency levels throughout the sessions/tests (Kawauchi, 2005). This measure, which is considered as an indication that some sort of monitoring is taking place, was calculated as follows: the number of repairs divided by the speaking time (Levelt, 1989; Michel, 2017; Tavakoli & Skehan, 2005). **Table 8** below shows examples for the CAF measures used in this study to assess L2 proficiency:

Table 8: Measures used for assessing CAF in the current study

	Type of Measure	Description	Example
Complexity Measures	Syntactic Complexity	Number of clauses per AS-unit (Yuan & Ellis, 2003). Length of AS-units (i.e., Number of words per AS-unit) (Kawauchi, 2005).	Anas: Presumably the concert would be an instant satisfaction but a pain on the long run (1 AS-unit, two clauses, 15 words)
	Complex grammatical structures	Use of relative clauses and other complex structures per task (Kawauchi, 2005)	Majd: the second patient <u>who </u>is an accomplished violinist and a 12 yo child should receive the heart because she's literally a child and she still hasn't lived that much

	Syntactic Variety	Total number of different grammatical verb forms used in the task in terms of tense, voice, and modality (Yuan & Ellis, 2003)	Majd: Well, I **think** the teacher **should receive** it because first he **has** two children that he **has** to take care of and **he is basically teaching** generations so he'**s** a treasure for sure In this example, the learner produced different grammatical verb forms: Tense = simple present/ present progressive And 1 instance of modality ("should receive")
	Lexical variety/richness Lexical accuracy/appropriacy	The total number of different words occurring in a text or utterance was divided by the total number of words (Vercellotti, 2017; Yuan & Ellis, 2003). Learners might use a variety of words during a task; however, the words might not be appropriate for the task context (in terms of meaning).	**e.g.,** In the bottom right corner, there's a **garbage can,** which is full. Instead of saying: there's a **trash can/bin** which is full
Accuracy Measures	Grammatical Accuracy	Number of error-free clauses per total number of clauses per task Number of errors as compared to the total number of words produced per task	e.g., If this **happen** to me, I will prove myself. In this example, we have two clauses; only the second one is error-free whereas the first one has a subject-verb agreement problem.

Fluency Measures	mean length of unfilled/silent pauses	total silence greater than 0.25	
	mean length of filled pauses	learners' using (erm, err)	
	Speech rate	mean number of meaningful syllables divided by speaking time, excluding, repetitions, false-starts. mean number of false-starts, reformulations per minute	
	Repair measures	repetitions, false-starts. mean number of false-starts, and reformulations per minute	

I now provide a description of the qualitative data collection sources and measures used to investigate the remaining research questions which concern effects of planning/no planning on learners' L2 performance, learners' planning strategies, as well as participants' attitudes and perceptions towards the whole online experience.

3.3.8.2 Qualitative Data Analysis

As discussed earlier (see **section 3.3.8**), given the time/word limit and due to the huge amount of data gathered from learners' planning notes, chat logs, stimulated recall interviews, as well as reflective reports, my supervisor and I agreed that data from only a small sample of the learners (i.e., four case studies) would be qualitatively analysed. Dörnyei (2007) contends that due to subjectivity issues, qualitative interpretations of the data can sometimes be problematic. That is to say, rather than objectively analysing the existing data, the researcher might be tempted to choose to show only what confirms their hypothesis/beliefs. Having analysed learners' chat logs and tests, I found that some participants achieved greater/fewer gains in

terms of the complexity, accuracy, and fluency of their L2 production. Although this could be traced back to the different treatments they received during the online sessions, it could also be due to other factors, including individual differences, rapport between the SLs and their NSs/STs partners, etc. Therefore, when choosing data for qualitative analysis, my aim was to find participants who represent all these cases so that more detailed information could be obtained about each case. It was hoped that this procedure would reduce subjectivity issues in the current study. According to Yin (2014), "a case study allows investigators to focus on a case and retain a holistic and real-world perspective" such as in scrutinising small group behavior (p.4). Therefore, it was believed that such an approach would help me make good use of the time available to thoroughly examine the cases that showed distinctive behaviour in the data collection stage to answer the third and fourth research questions via means of stimulated recall interviews. This includes understanding the rationale behind learners' planning choices and their production of more/fewer complex structures, errors, and pauses throughout the sessions. It also involves examining the rapport or lack of rapport between the learner and their partner, and finally looking into ways in which external factors, such as anxiety/motivation level, have impacted learners' overall performance. Answering these questions, which were often neglected by previous research that mostly focused on quantitative analysis, has the potential to reveal why some learners showed greater progress than others regardless of their group condition (i.e., planning or control group). An overview of each case study will be provided below.

- Majd, an IT student, was very motivated to practice his L2 as he was preparing to go to Poland to attend an IT course in English. During the sessions, he was assigned to work in the treatment group and had to interact online with a NS. Further analysis of his online performance disclosed that Majd was benefitting from the feedback delivered by

his partner, Rosy (which was mostly explicit, especially in the text chatting mode) and was producing more complex, accurate, and fluent output as the sessions proceeded.

- Lara, who had been studying French by the time this study was conducted, was also paired with a NS and like Majd and Rosy, they worked together under the guided planning condition (which was considered as the treatment group in the current study). It is worth noting that Lara and her partner, Kate, were getting on really well and developed a very good rapport as they were working on the tasks. Qualitative analysis of the chat logs showed how Lara was gradually progressing throughout the sessions, becoming less anxious and hence more confident about her language use.

- Similar to Majd, Ameen was also preparing to go to Poland to attend an IT course in English. Given the fact that he was working in the treatment group, Ameen was not showing much progress in terms of his language use compared with Majd and Lara. Having analysed his chat logs as well as responses during the recall interviews, a lack of rapport between Ameen and his partner, Rami (a Syrian English teacher who is currently living in the UK) was very evident throughout the online sessions. It is worth noting though that during the interviews, Ameen seemed a bit shy and reticent about elaborating on his language even when using his L1. Thus, the basis on which Ameen was selected as a case study was to examine whether non-linguistic factors like lack of rapport between the interlocutors and/or individual differences might have occasionally hindered learners' L2 development.

- Rima, on the other hand, was assigned to work in the control group and had to interact online with Rana (also a Syrian English teacher). Rima only received instructions related to task goal/content prior to the online performance and was asked to start the session immediately. Qualitative data from her chat logs disclosed a change in her online behaviour; Rima was getting more motivated about improving her English as the

sessions were proceeding and was asking questions about her language use, especially with regard to her word choice and sentence structure. In some sessions, Rima was the one who was doing most of the talking and leading the discussion with her partner. Throughout the sessions, Rima was showing some progress and was writing/speaking more compared with the previous sessions. Hence, the rationale for choosing Rima for the qualitative analysis was twofold: firstly, to allow comparisons with the other three case studies who were performing the online tasks under the guided planning condition; and secondly, to substantiate the quantitative findings which will be discussed in the next chapter and explicate the progress that was achieved sometimes by the control group, in terms of the complexity, accuracy, and fluency of their production during the subsequent post-/delayed tests.

The following sections will demonstrate data analysis procedures for each of the employed data sources.

3.3.8.2.1 Planning notes:

In an attempt to identify the strategies SLs followed to plan for tasks that varied in type and complexity, the qualitative data involved a "content analysis of emergent themes" (Ortega, 2005, p. 83). The notes taken by the SLs were coded based on the aspects of language they focused on (i.e., grammar, meaning/content, vocabulary, and spelling). Then the learners were invited to comment on these notes as they were watching the video recordings, explain what was going on, and justify their choices. **Table 9** below details and illustrates all the codes used to analyse learners' planning notes.

Table 9: Codes used for the analysis of SLs' planning notes

Language Aspects	Definition	Example
Grammar	Instances where the learners attended to the grammatical aspects of their production (e.g., sentence structure, tense, articles, prepositions).	Majd: The main problem that Linda has, is to find the time between her responsibilities. *Majd then edited this sentence to read as follows:* The main problem is that Lind has to find time for her different responsibilities.
Meaning	Instances where the learners wrote ideas related to the content of the task	Lara: On the shelves, sound system
Vocabulary	Instances where the learners listed single words to refer to particular items they saw in a picture, for example.	Ameen:. bed Chairs Candles Pillow
Spelling	Instances where the learners kept editing a word till they figured out the right spelling	Rima: carbage (**edited to become** garbage)
pronunciation	Instances where the learners used an online dictionary to check how a particular word is pronounced	N/A

I also compared the overall strategies followed by the participants to detect any patterns or differences due to working on tasks that differed in type and complexity throughout the sessions. That is, I looked for differences in the content/focus of their notes as the tasks progressed.

3.3.8.2.2 Chat logs

Only the online transcripts of four SLs were coded for instances of CAF (as discussed above) (e.g., pauses, self-correction, and implementation of complex structures) as well as instances of corrective feedback. I began by drawing up a start list of categories and subcategories of codes. Instances of feedback were operationalised as episodes in which SLs produced non-target-like forms and as a result received either explicit or implicit feedback from their NS/ST partners. Other instances of feedback involved the NS/ST commenting on the sentence structure produced by the SL and recommending revision using more complex structures (e.g., relative clauses). Explicit feedback includes the provision of direct or metalinguistic information about the nature of the error. Implicit feedback, on the other hand, can be divided into two components, recasts and modification devices (Long, 1996). Recasts refer to the target-like reformulations through which L2 users indirectly corrected their partners' errors without essentially breaking the flow of the conversation. Unlike a recast, modification devices vary in their explicitness and do not contain a full reformulation. Rather, they require learners attempt self-repair or output modification and hence promote deeper processing of already internalized L2 forms. These include: clarification requests (e.g., what do you mean?); comprehension checks (e.g., did you understand?); elicitation (can you reword this sentence?) and confirmation checks (e.g., do you mean X?) (see Long, 1996). These devices were also employed sometimes by NSs/STs to give SLs opportunities to clarify their utterances and pick up new forms. All turns containing instances of feedback were coded with respect to the aspects of language that triggered them (e.g., lexical, grammatical, spelling, content, etc.) (See **Table 10**).

Table 10: Coding Categories of the type of feedback provided by NSs/STs

Category	Level	Examples
Language focus	Lexical Grammatical	SL: it just has a curved and straight lines on it NS: we would say 'curvy lines' or wavy lines' SL: and there's a chair in front on the computer desk NS: you mean, 'in front of'?
Type of feedback	Explicit Or Implicit (Recast) (Confirmation Check)	SL: Should I start telling you what I'm seeing in the picture? NS: yes please. Also, you would say 'what I see in the picture'. Present simple instead of present continuous. SL: on the left upper corner, NS: in the upper left corner? SL: in the down left corner, NS: you mean, 'in the bottom left corner'

3.3.8.2.3 Stimulated recall interviews:

Since using interviews, viewed as the most important sources of case study evidence, was designed to provide more details about the quantitative results, the chosen case studies were the ones that contributed the most to the quantitative dataset (e.g., featuring learners who achieved higher CAF scores, were more motivated than others) (Yin, 2014). Recordings were transcribed and transformed into texts. Then these texts were coded in two stages. In the initial stage, texts were read and information related to my research was labelled. Then in the second stage, labels from the four participants' interviews were identified, analysed and clustered together under broader labels, as in the example below:

"I felt intimidated most of the time during the online session **(intimidation)**. The fact that my partner is a native speaker of English made me think that the conversation should be very formal **(L2 background)**. She corrected me whenever I made an error

or wrote an informal word. To be honest, this stressed me out during the session **(anxiety)** which was why I tried be more careful about my language **(focus on form)** and I did not want to make more errors **(focus on accuracy)**", Majd commented.

Another example was when the learners were asked to comment on one of the instances where NSs/STs provided feedback or commented on learners' language use to see whether there was an uptake of that feedback. If the SL's recollection of the chatting segments focused on the language forms and/or the lexical items, it was considered as evidence of reported uptake. Whereas, if the SL reported that s/he could not recall what was going on at a particular time during the interaction or talked mainly about the task's content or some other aspect of task completion, it was classed as no reported uptake. Additionally, stimulated recall interviews were analysed to identify participants' attitudes and perceptions towards the online experience. The final code list was then produced[20].

In order to ensure the reliability of the codes list, an inter-rater reliability procedure was used. I asked an independent rater, a PhD student studying TESOL at the University of Sheffield, to recode 40% of the data. The data included samples of learners' performance during the tests/the online sessions as well as learners' planning notes, interview scripts, and reflective reports. An agreement rate of 84% was obtained. Then, we started to compare and debate problematic areas, especially data from chat episodes which were subject to different interpretations (e.g., when the participant went off topic, instances where there was uncertainty whether the feedback was implicit or explicit, or whether the learners were focusing on form/meaning). New codes, definitions, and examples were created as necessary till agreement on each code was achieved. What follows are examples that demonstrate some of the problematic codes and how they were modified.

---

[20] see appendix G

Example 1. Majd: in the **<u>down</u>** left corner, there's a table with two candles on it.

This sentence was firstly coded as an indication that the learner was focusing primarily on meaning, given the error he made (writing '***down*** left corner' instead of '***bottom*** left corner'). During the interview, however, the learner mentioned that he was aware that he used 'down' incorrectly in this sentence, but he was just attempting to deliver the meaning to his partner, so she would be able to correct him. Hence, the independent rater advised that it would be more accurate to code this as *focusing on meaning and form simultaneously* since the learner acknowledged that his use of the wrong preposition was intentional after he failed to figure out the right one.

Example 2. Majd: there's a chair in front **<u>on</u>** the computer desk.

Another instance of disagreement was about whether or not to code learners' typos as errors, as in Example 2, especially when the learners reported these as typos during the interviews. The independent coder proposed that these would still be coded as spelling errors.

Example 3. Lara: tea-towel?

what does that mean?

Kate: Yes, that's what we call it in the UK

Do you use tea-towels in Syria?

I know in some countries they don't.

Example 3 was first coded as an instance in which the learner went 'off topic', but then based on Kate's comment in the recall session:

"At this point, I was not sure whether or not Lara understood what the word 'tea-towel' means so I asked her if they use tea-towels in Syria, but then I thought it sounds like a silly question. Lara might have thought ' what does this girl think we are in Syria', so I did follow it up by saying 'I know in some countries they don't.'"

the independent coder suggested that it would be better to code it as 'cultural reference' that was one of the strategies used to check whether or not the SL understood the meaning of a particular lexical item. All the abovementioned instances of disagreement were modified accordingly.

3.4 Summary of the Chapter

This chapter delineates the research design of the current study. The methods employed for data collection and data analysis were also justified. To conclude, there were many limitations and challenges to the project that I could not have foreseen. Neither could I have predicted how much time these issues would take to resolve. These included selecting appropriate tasks in terms of type, content, and complexity for all the online sessions to serve the purposes of this project. Notwithstanding that task selection sounds like a fairly routine part of planning this study, the fact that most of the tasks used in previous CAF research were based on monologic production made most of them unsuitable to be replicated in the current study; hence, it was quite challenging to come up with new tasks. Participants not following the instructions (e.g., providing feedback on learners' errors, inducing the use of relative clauses) also caused a lot of problems since I had to revise the CAF measures that were employed to operationalise L2 development. The quality of the Internet connection in Syria made the process of data collection even harder as I had to reschedule a number of sessions, so that the SLs had a better connection. Recruiting participants who would be willing to take part in a longitudinal study was another problem; many participants dropped out of the study as they could not fully commit to the online sessions that lasted for five months. It is worthwhile stating that the decision to collect this much data over a considerable period of time was influenced by the gaps found in previous research and the desire to provide a more methodologically rigorous design. I believed that the longitudinal nature of the study would help enable a more methodologically robust study to measure the participants' learning.

Having discussed the pilot study as well as all the amendments made to the methodology of the main study, the next chapter will display and discuss the quantitative/qualitative findings relating to the research questions addressed in this project.

4. <u>Results & Discussion (1)</u>

The results and discussion of the present study will be split into two chapters. The first chapter will present and discuss the quantitative data related to research question 1 and research question 2, whilst the following chapter will address the qualitative findings for research question 3, 4, and 5.

<u>Part 1: Quantitative Findings</u>

This chapter displays the quantitative results from this study in the sequential order of the research questions. Learners' production during the tests (pre-, post-, and delayed) were utilized as the main data sources to introduce the current findings. The first part of this chapter tackles the effects of manipulating task complexity and task type during the online sessions on learners' L2 production across the subsequent tests. In particular, I intended to investigate whether manipulating the cognitive demands or the cognitive conditions of tasks over time would improve learners' L2 proficiency and help them access multiple features of production (i.e., complexity, accuracy, and fluency, or CAF) simultaneously. Learners' production in the pre-, post-, and delayed tests, as divided per group, will therefore be compared to detect the changes, if any, due to the different treatments they received during the online sessions.

As previously mentioned, task complexity was manipulated by means of planning conditions and hence twenty SLs interacted online with more proficient L2 interlocutors and were assigned to either the experimental or the control group. In the experimental group, SLs were provided with explicit instructions to focus on certain linguistic forms (e.g., articles, relative clauses, prepositions) and ten minutes prior to the commencement of the online session to take notes and prepare for their online performance. In contrast, in the control group, learners received general instructions to achieve the task goal, and were asked to start working on the

task with their partners immediately. Statistical analyses were carried out using SPSS and hence descriptive statistics for each of the CAF variables utilized to operationalise L2 development in two different modes of online interaction will be provided in the following sections.

4.1 RQ1: Does manipulating task complexity during online sessions have an impact on learners' L2 development?

During the first step of data analysis, independent samples t-tests were conducted to detect any significant differences between the treatment group and the control group in terms of the complexity and accuracy of their L2 production prior to the commencement of the online sessions. As far as the complexity of learners' production is concerned, this was measured via means of syntactic complexity (i.e., ratio of clauses per AS-unit, average length of AS-unit, use of complex grammatical structures), syntactic variety (i.e., total number of different grammatical verb forms used in the task, in terms of tense, voice, and modality), lexical variety as well as lexical appropriacy.

4.1.1 Results of SLs' written production during the pre-test

No differences were found in SLs' pre-test results in terms of the complexity and accuracy of their written production. For example, although the descriptive statistics presented in **table 11** for the first complexity measure showed that the ratio of clauses per AS-units for the control group (M= 2.21, SD= 0.54) was slightly higher than that of the treatment group (M= 2.10, SD= 0.47), this difference was not significant (t (18) = .928, p = .366). Similar results were found for all the other complexity and accuracy measures during the written pre-test. This confirms that all the SLs were considered homogenous in terms of their written L2 proficiency at the beginning of this study (**see graph 1**), and hence it would be easier to determine whether or not the treatment received by the guided planning group was more effective in terms of L2 development.

Table 11: descriptive statistics (1) for CAF measures in the written pre-test

CAF measures	Groups N= 10	*M*	*SD*	*Minimum*	*Maximum*
		Pre-test			
Ratio of clauses per AS-unit	TG	2.10	0.47	1.44	3.00
	CG	2.21	0.54		
Average length of AS-unit	TG	13.49	3.50	9.22	23.5
	CG	15.75	4.29		
Complex grammatical structures	TG	0.30	0.67	0	3
	CG	0.50	0.97		
Syntactic variety	TG	4.10	1.37	2	7
	CG	3.30	1.16		
Lexical variety	TG	66.61	17.36	23.24	100.0
	CG	58.48	13.70		
Lexical appropriacy	TG	1.00	1.05	0	3
	CG	0.60	1.07		
Error-free clauses	TG	57.61	14.90	25.00	82.53
	CG	55.60	16.68		
Number of errors per test	TG	10.27	3.70	2.89	20.1
	CG	10.36	5.41		
Number of words per task	TG	12.84	3.71	6.30	19.0
	CG	9.69	2.72		

Graph 1: CAF results per group during the written pre-test
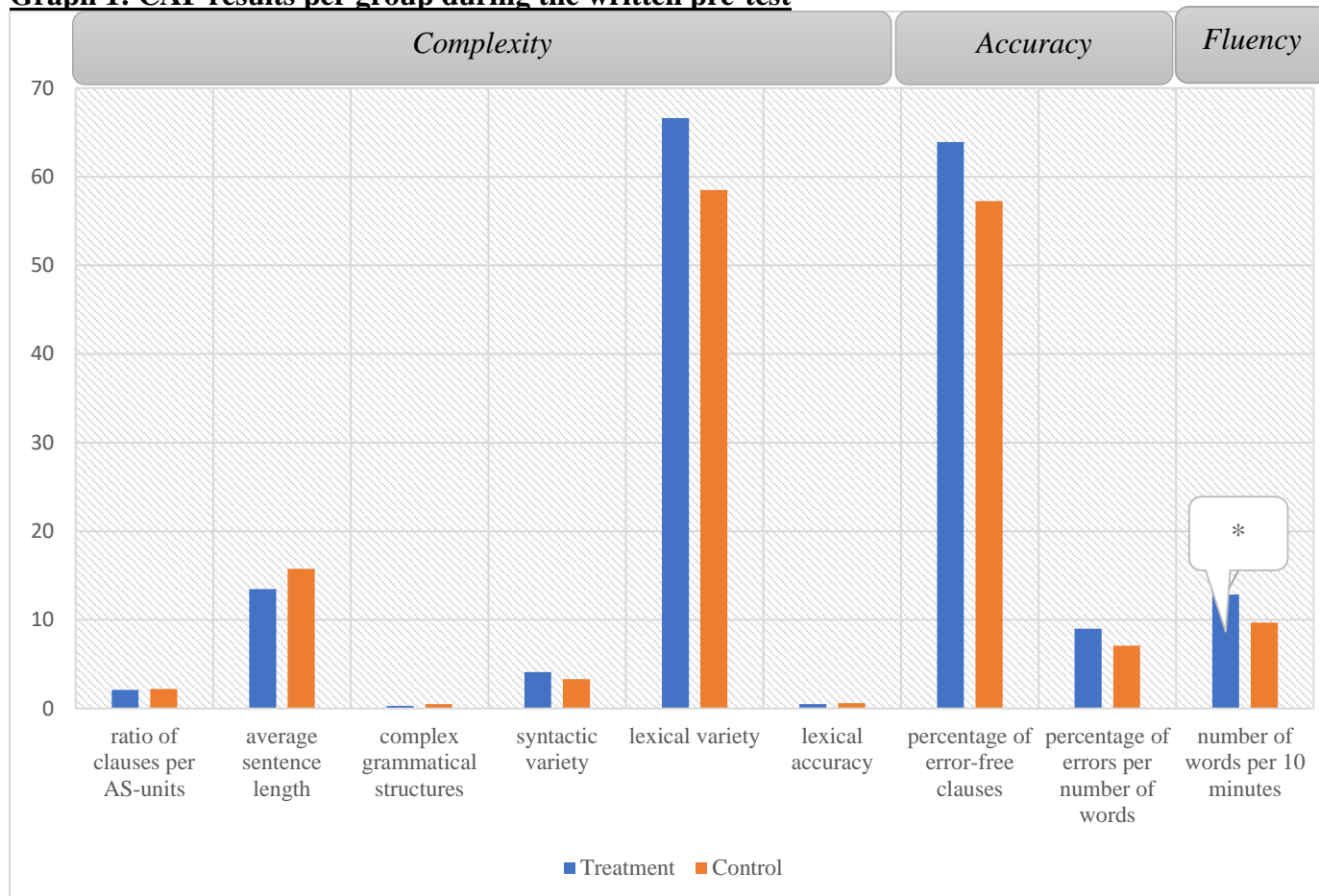
As shown in **graph 1** above, the number of words per minute (excluding repetitions, deletions) was the only measure that I used to assess the fluency of SLs' written production during the pre-, post-, and delayed tests (Allaw & McDonough, 2019; Asgarikia, 2014; Ellis & Yuan, 2004). This was mainly due to the poor Internet connection and the lack of access to keystroke logging software. Therefore, it was not possible to measure learners' fluency during the written tests. The number of words produced per minutes, however, was considered as an indication of development in learners' L2 fluency. **Table 11** also shows that there was a difference between the treatment group and the control group in terms of the length of their texts as compared to the time they spent on the task during the pre-test which was found to be significant (*t* (18) =2.15, *p*=.04). To put it another way, learners in the treatment group produced more words than the control group given the time they took to finish their texts in the

pre-test. Although this was considered as an indication of fluency in the current study, we cannot safely assume that learners in the treatment group were more fluent than those in the control group; the latter could have produced more words, yet revised their texts a lot before submitting their final draft. Additionally, the typing skills of some learners could sometimes be a hinderance to writing long texts, especially when restricted to a short period of time.

4.1.2 Results of SLs' written production during the immediate and delayed post-tests:

As for the second step of data analysis, independent samples t-tests were conducted for each of the CAF measures during the immediate and delayed written post-tests. The statistical tests aimed to compare the performance of both groups and evaluate the impact of the guided planning as well as the explicit feedback/instructions on the treatment group's L2 written proficiency. As the descriptive statistics in **Table 12 & 13** show, overall and for most of the CAF measures (except for lexical variety for delayed test 2, and lexical accuracy measures), there was a difference in the mean scores between the treatment group and the control group in favour of the former. To put it simply, there was a trend of the treatment group performing better than the control group across the immediate and delayed post-tests. This indicates that there was some effect of the guided planning and the feedback provided for the treatment group. The following paragraphs will introduce and elaborate on the findings related to each of the dependent variables employed to analyse learners' written tests: independent samples t-tests showed that SLs achieved significant gains in the current study in terms of the complexity and fluency of their written production. However, no statistically significant results were captured with regard to the accuracy measures.

Table 12: Descriptive statistics for CAF measures: Treatment group & Control Group

CAF measures	Group N=10	Post-test 1		Delayed test 1		Post-test 2		Delayed test 2	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Ratio of clauses per AS-unit	TG	2.62	.48	2.52	.43	2.79	.44	2.50	.35
	CG	1.96	.40	1.71	.32	1.96	.44	1.76	.36
Average length of AS-unit	TG	15.95	3.69	15.39	2.73	18.60	4.19	15.97	3.0
	CG	11.30	2.44	11.18	1.29	12.87	2.75	12.06	3.77
Complex grammatical structures	TG	.70	.675	.80	1.03	2.10	2.02	.80	.919
	CG	.50	.707	.50	.527	.50	.527	.30	.483
Syntactic variety	TG	3.50	1.08	4.40	.84	4.0	1.24	4.60	.69
	CG	3.40	.96	3.60	1.17	3.70	.82	3.00	1.24
Lexical variety	TG	68.9	26.6	60.88	13.92	54.93	15.47	57.35	6.75
	CG	49.2	8.94	44.09	8.07	49.29	9.07	59.34	15.94
Lexical appropriacy	TG	1.50	1.26	.80	1.22	1.40	.843	.90	.994
	CG	0.70	1.33	.50	.707	.60	.843	.10	.316
Error-free clauses	TG	66.02	14.24	74.09	13.31	47.33	17.29	66.97	8.63
	CG	57.36	20.49	61.28	24.43	35.83	22.87	53.35	19.51
Number of errors per test	TG	9.38	4.87	6.21	3.11	14.29	5.25	8.95	3.99
	CG	11.16	5.61	8.66	5.14	16.61	7.16	12.24	7.27
Number of words per task	TG	11.66	3.25	14.67	3.79	15.48	3.79	15.32	3.73
	CG	10.97	2.31	11.42	2.28	11.73	2.51	12.95	2.90

Note. TG=treatment group, CG=control group, M=mean, SD=standard deviation, N=number of participants per group

Table 13: Descriptive statistics for CAF measures: Treatment group & Control

	Post-test 1		Delayed 1		Post-test2		Delayed 2	
	min	*max*	*min*	*max*	*min*	*max*	*min*	*max*
Ratio of clauses per AS-unit	1.44	3.25	1.14	3.37	1.27	3.66	1.16	3.14
Average length of AS-unit	8.00	20.5	9.62	19.4	9.00	27.5	7.08	20.8
Complex grammatical structures	0	2	0	3	0	6	0	2
Syntactic variety	2	5	2	6	2	6	1	6
Lexical variety	37.9	132.5	34.1	80.4	36.6	84.9	36.8	83.0
Lexical appropriacy	0	4	0	4	0	3	0	3
Error-free clauses	12.5	92.3	11.7	100	3.76	72.0	21.7	79.4
Number of errors per test	1.21	18.51	.00	17.3	5.88	26.1	4.47	27.5
Number of words per time spent on task	7.60	18.90	6.90	19.3	8.80	22.0	8.50	21.1

4.1.2.1 Complexity measure 1: ratio of clauses per AS-units

This was calculated by dividing the number of clauses by the overall number of AS-units the learners produced per test. Thus, the formula used in determining syntactic complexity was:

Number of clauses
Number of AS-units

Independent samples t-tests revealed that the treatment group had higher mean scores, and was producing more clauses per AS-unit than the control group across all the immediate and delayed tests. All the results were found to be statistically significant between the groups: immediate posttest 1 (t (18)=3.24, p=.005), delayed test 1 (t (18)= 4.67, p=.000), immediate posttest 2 (t (18)= 4.14, p=.001), and delayed test 2 (t (18)= 4.62), p=.000) (see **graph 2**). This

probably could be traced back to the feedback they received from their partners regarding their language use; NSs/STs were trying to push the learners to produce more extended and comprehensible utterances. Hence, maybe as the study proceeded, the learners started to benefit from their partners' comments and consequently began to produce more elaborated language compared with the control group. In addition, the fact that the treatment group had to plan their L2 performance prior to each of the online sessions could have had an effect on the learners' language structure elaboration and could have contributed to learners' production of more complex AS-units during the tests. To put it another way, SLs, especially those who allocated their attention primarily to focus on meaning while planning, could have automatized their L2 knowledge of particular structures and chunks that could be used in different L2 contexts and hence they employed these whenever they found this appropriate throughout the tests.

Graph 2: ratio of clauses per AS-unit
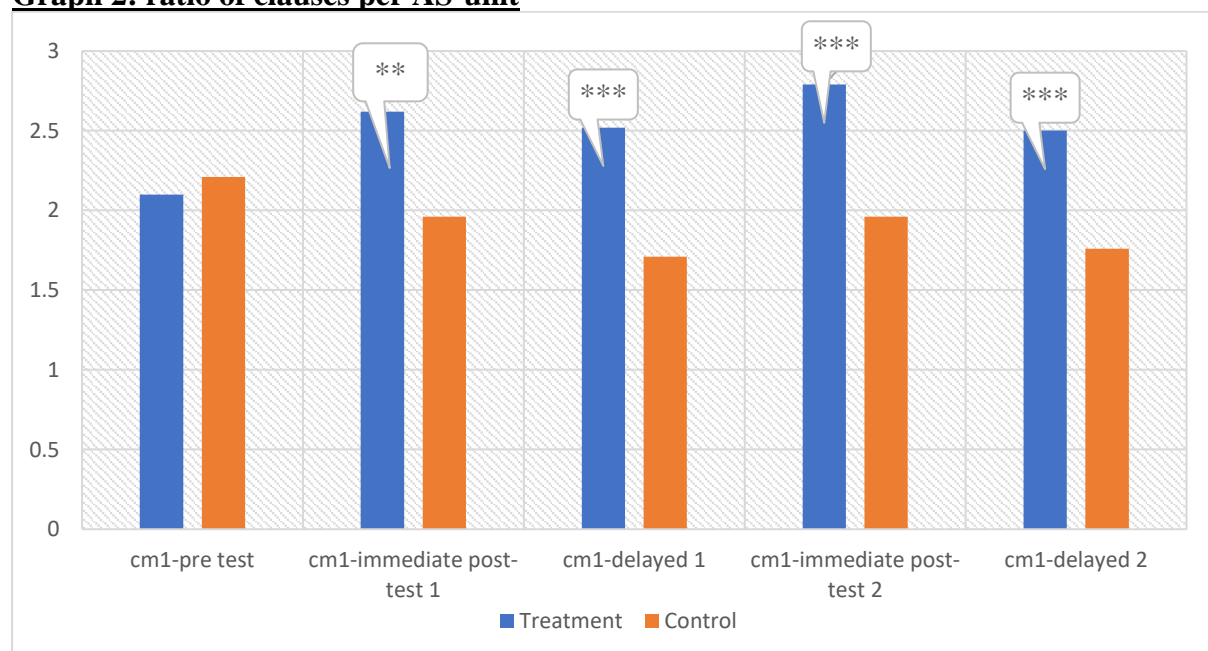
Note. cm1= complexity measure 1, ** = $p \leq$ .01, ***= $p \leq$ .001

4.1.2.2 Complexity measure 2: average length of AS-units

Findings obtained from independent samples t-tests to capture the difference between the two groups regarding the number of words they produced per AS-unit per test were very consistent with those for the previous measure. Overall, the experimental group had higher mean scores

(i.e., produced longer AS-units) than the control group. This difference was statistically significant throughout all the immediate and delayed tests: immediate posttest 1 (t (18)= 3.31, p=.004), delayed test 1 (t (12.85)= 4.40, p=.001), immediate posttest 2 (t (18)= 3.60, p=.002), and delayed test 2 (t (18)= 2.56, p=.01) (see **graph 3**).

Graph 3: length of AS-units per group per test
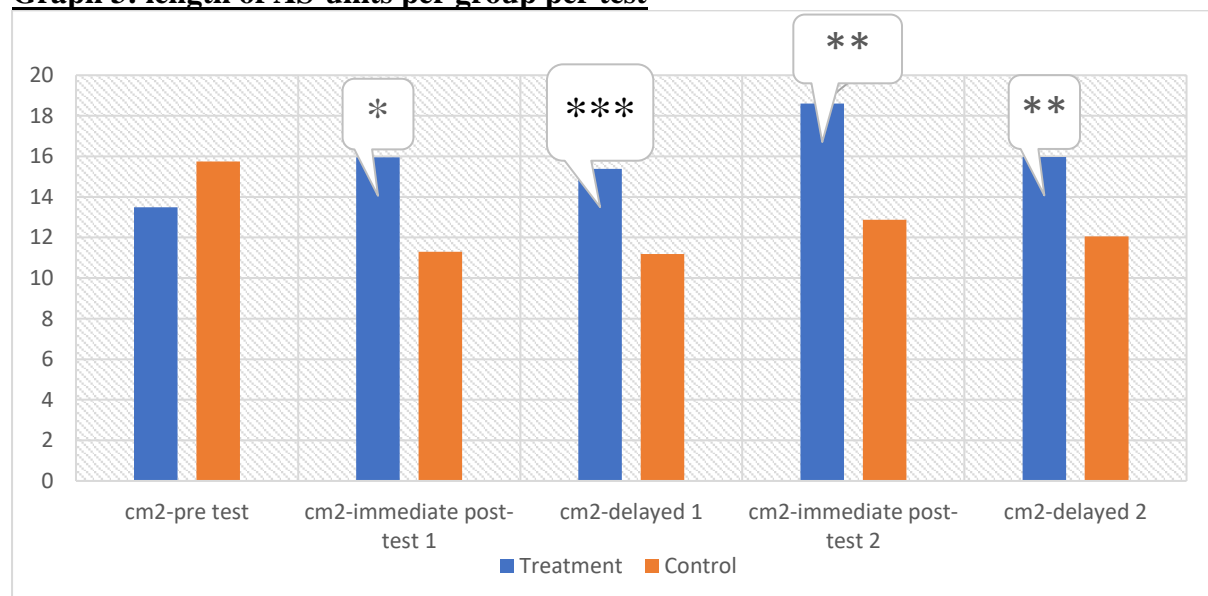
Note. **cm2= complexity measure 2, **= $p \le$ .01, ***= $p \le$ .001**

Again, the longitudinal explicit treatment and the planning time received by the treatment group probably helped them attain greater and steadier gains than the control group during the tests.

4.1.2.3 Complexity measure 3: production of complex grammatical structures

Apart from the pre-test, there was a tendency of the treatment group to produce more complex grammatical structures (i.e., relative clauses) than the control group throughout the tests (see **table 14**). This result indicates that the combined effects of guided planning, task complexity, as well as the feedback provided to the treatment group with regard to the implementation of relative clauses did have a positive impact on learners' production of these forms.

Table 14: learners' production of complex grammatical structures per test

	groups	N	M	SD	t	df	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Written pre-test/complexity measure 3	planning	10	.30	.675	.535	18	.986	.586
	control	10	.50	.972			.993	.593
Written immediate post-test 1/ complexity measure 3	planning	10	.70	.675	.647	18	.449	.849
	control	10	.50	.707			.450	.850
Written delayed test 1/ complexity measure 3	planning	10	.80	1.033	.818	18	.470	1.07
	control	10	.50	.527			.490	1.09
Written immediate post-test 2/ complexity measure 3	planning	10	2.10	2.025	2.41*	10.21	.210	2.99
	control	10	.50	.527			.130	3.07
written delayed test 2/ complexity measure 3	Planning	10	.80	.919	1.52	13.62	.190	1.19
	control	10	.30	.483			.206	1.20

Note. * = p ≤ .05

However, no statistically significant difference was revealed between the groups regarding learners' implementation of these complex structures, except in immediate posttest 2 (t (10.21) =2.41, p=.03). This was possibly due to the direct effects of the voice chat sessions where the treatment group was *listening* to the language produced by their partners and *receiving* somehow more instant feedback regarding the incorporation of these complex forms into their L2 production.

> **E.g.,** Majd: there is an ashtray but there is no cigarette
>
> Rosy: ok, so you see an ashtray <u>which doesn't have a cigarette in it</u>

In other words, the planning group might have benefitted from the immediacy of turn-taking, a feature that does not often exist in text chatting due to the delayed interaction. That is, by the time the learners got their partner's message in the written mode, they might have moved on to

talk about something else and hence did not notice the feedback provided in the previous turns. However, the non-significant results between the groups across the other tests could be traced back to different reasons. Firstly, learners working under the treatment group might have found these forms (i.e., relative clauses) rather complex and did not feel confident enough to integrate them into their L2 production during the tests. To put it another way, though SLs comprehended these forms as they saw them on the screen and reported that sometimes they tried to imitate their partners during the online sessions by producing similar structures (see **section 5.4.1.1**), they felt less comfortable and confident using relative clauses when working individually. Thus, understanding particularly complex grammatical forms while reading or listening (i.e., relative clauses) does not necessarily mean having the ability to incorporate these forms into their production when speaking or writing. NS/STs, working with the planning group, were frequently requested to explicitly draw learners' attention to particular grammatical forms. However, and particularly during the text-based chat sessions (probably due to the lack of rapport and F2F interaction with the learners), they preferred to provide indirect feedback, which was not always understood by the learners as such.

4.1.2.4 <u>Complexity measure 4: Syntactic variety</u>

Syntactic variety was measured via means of the total number of different grammatical verb forms used per task, including tense, voice, and modality. There was a trend for the treatment group to produce more variety of verb forms than the control group in the current research (see **table 15**), yet no significant difference was found in the mean scores between the groups, except for delayed test 2 (t (18) =3.53, p=.002).

Table 15: variety of verb forms per test

	groups	N	Mean	SD	t	df	95% Confidence Interval of the Difference	
							Lower	Upper
Written pre-test/complexity measure 4	planning	10	4.10	1.370	1.40	18	.393	1.99
	control	10	3.30	1.160			.395	1.99
Written immediate post-test 1/ complexity measure 4	planning	10	3.50	1.080	.218	18	.863	1.06
	control	10	3.40	.966			.864	1.06
Written delayed test 1/ complexity measure 4	planning	10	4.40	.843	1.75	18	.160	1.76
	control	10	3.60	1.174			.167	1.76
Written immediate post-test 2/ complexity measure 4	planning	10	4.00	1.247	.635	18	.693	1.29
	control	10	3.70	.823			.704	1.30
Written delayed test 2/ complexity measure 4	planning	10	4.60	.699	3.53**	18	.650	2.55
	control	10	3.00	1.247			.631	2.56

Note. ** = p < .01*

One plausible reason for the lack of significant progress achieved by both groups regarding this measure could be the nature of the task type the learners worked on during the tests (story-telling tasks based on a series of incidents), and hence the learners were mostly using simple tenses (past or present). Additionally, the varied mean scores attained by the learners across the tests might be a rational outcome of the complexity or otherwise of the task's content as perceived by the learners. The treatment group in particular could have benefited from their partners' feedback in the online sessions; and therefore addressed the tasks that had a simple content using simple language forms to fulfil the task's goal, whereas more cognitively demanding tasks in terms of content and structure perhaps required the use of a different, wider

variety of verb forms. The opposite could have happened, however, for the control group, who only received instructions that tackled the task goal during the online sessions: NSs/STs working with the control group were not instructed to focus on learners' correct use of particular forms/tenses. Hence, possibly due to the lack of language instructions, the control group did not attempt to address the more complex tasks by employing a wider variety of structures and verb forms.

The significant results between the groups in delayed test 2 could be an indication that the treatment group was making gradual progress and utilising a greater variety of verb forms in their output as the treatment progressed. This slow progress was indeed influenced by the duration of the online sessions which was relatively short, and thus longer and more frequent L2 exposure could have led to a more noticeable difference between the groups. As reported in previous CAF research, the fact that learners' attentional resources were directed towards different aspects of L2 production (i.e., focusing on form, meaning, spelling, and word choice) during the online interaction might have contributed to decelerating learners' progress as well (Ellis, 1994; Larsen-Freeman; 2006; Robinson, 2001; Skehan & Foster, 1999; Yuan & Ellis, 2003).

4.1.2.5 Complexity measure 5: lexical variety

Lexical variety was measured via means of MTLD (the measure of textual lexical diversity), which is calculated by dividing the overall number of different words that occur in a text by the number of its tokens (i.e., running words). There was a significant difference between the treatment group and the control group only during immediate posttest 1 (t (18)= 2.21, p=.04) and delayed test 1 (t (14.4) = 3.29, p=.005) (see **graph 4**). The treatment group, which was instructed to try to benefit from their partner's language use, produced a greater variety of lexical items during the tests. This might be also due to the nature of text-based chat as learners

were seeing their production and that of their partner on the screen and hence it may have been easier for them to learn new vocabulary and lexical chunks and reuse them during the following tests. This could also explain the non-significant results between the groups in the following tests (i.e., immediate post-test 2 and delayed test 2) after the intervention of the voice chat sessions, where learners were just listening (without seeing the conversation on the screen), and probably not understanding all the words produced by their more competent L2 partners. Accordingly, it would be highly unlikely for the learners to integrate these words into their own utterances. Other possible explanations for the non-significant results between the groups in the last couple of tests could be due to the tasks' complexity (recall that tasks were sequenced from simple to complex based on the variety of vocabulary/structures needed to create a narrative). Therefore, by the time learners got to the last two, most complex tasks, they may have found them too complex, which in turn could have affected their production of a variety of lexical items. Learners' individual differences (i.e., how they understood the scenario of each story) might have also influenced learners' vocabulary choices during the tests. That is, learners had different interpretations and scenarios with regard to the set of pictures they were given; some of which were more complex and forced the use of varied lexical items.

Graph 4: lexical variety per group per test
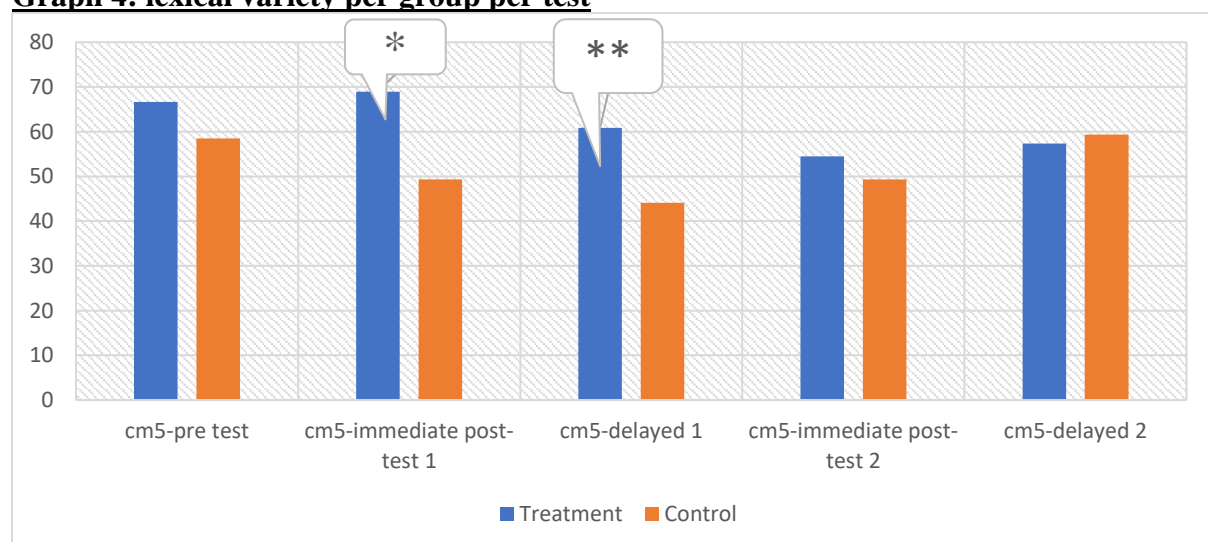
Note. **cm5= complexity measure 5, *= p≤ .05, **= p≤ .01**

132

Nevertheless, this does not necessarily mean that all the lexical items incorporated by the learners were appropriate and suitable for the given context. This issue will be further discussed in the following section when presenting the results for complexity measure 6.

4.1.2.6 Complexity measure 6: lexical appropriacy

This was measured by calculating the number of words produced by the learners that were inappropriate or inadequate in terms of meaning given the context of the task.

Table 16: lexical appropriacy per test

	groups	N	*Mean*	*SD*	*t*	*df*	95% Confidence interval of the Difference	
							Lower	*Upper*
Written pre-test/ complexity measure 6	planning	10	1.00	1.054	.840	18	.600	1.40
	control	10	.60	1.075			.600	1.40
Written immediate post-test 1/complexity measure 6	planning	10	1.50	1.269	1.37	18	.425	2.02
	control	10	.70	1.337			.425	2.02
Written delayed-test 1/ complexity measure 6	planning	10	.80	1.229	.669	18	.642	1.24
	control	10	.50	.707			.660	1.26
Written immediate post-test 2/ complexity measure 6	planning	10	1.40	.843	2.12*	18	.008	1.59
	control	10	.60	.843			.008	1.59
written delayed-test 2/ complexity measure 6	planning	10	.90	.994	2.42*	10.802	.107	1.49
	control	10	.10	.316			.072	1.52

Note. *= $p \leq .05$

There were differences in the mean scores (see **table 16**) between the treatment group and the control group throughout the tests, with the former producing more inappropriate lexical items than the latter. Yet the only significant difference was found during immediate posttest 2 and delayed test 2, i.e., after the voice chat intervention: (t (18) = 2.12, p=.04) and (t (10.80) = 2.42,

p=.03) respectively. The audio chat sessions could have affected the results of these tests; as the learners were not seeing their partners' turns on the screen, they might have confused the use of some lexical items. Another explanation for the treatment group using more inadequate lexical items than the control group (especially during immediate post-test 1 and immediate post-test 2), involves the different task type/contexts the learners were working on and the feedback they received in the online sessions. Thus, SLs were probably very eager to try the new forms that they read/heard during the interaction; however, these attempts were not always deemed to be successful (i.e., they did not know enough about these words to be able to use them in context). Accordingly, knowing the form and meaning of a word does not inevitably entail having the ability to use it appropriately.

The control group, on the other hand, who did not receive instructions or feedback on new forms could have preferred to produce mainly the words they were confident about, in order to minimize their errors and mistakes. Hence, producing fewer inappropriate words might not necessarily mean that the control group were showing progress with regard to this measure. It could rather mean that the learners were not willing to take any risks and were sticking to the forms they were familiar with. It is worth mentioning, however, that the control group were still exposed to the online interaction though they were not getting any feedback on their language use or any instructions to integrate particular forms. That is, the learners were still getting the opportunity to practise their English with more proficient L2 interlocutors, and hence one would expect to see some development in their overall performance as a result.

4.1.2.7 <u>Accuracy measure 1: percentage of error-free clauses per test</u>

Independent samples t-tests revealed that the treatment group made greater progress than the control group regarding accuracy measure 1 throughout the tests. In other words, the treatment group had a higher percentage of error-free clauses per test (see **table 17**). This might be traced back to guided instructions, planning time, and the feedback the experimental group received

during the online sessions related to their non-target-like forms and incorrect language use, as opposed to the control group who received neither. In fact, the control group were only asked to address the task goal, and hence, they were presumably primarily focusing on meaning rather than the formal aspects of their language. The difference between the groups in terms of error-free clauses was not found to be significant in any of the tests, however. Perhaps the short duration of treatment and the fact that some NSs/STs preferred to give implicit feedback on the learners' errors could explain the lack of significant results throughout the tests. Another explanation could be that the control group were only writing simple sentences and employing mainly simple language. This interpretation seems in line with the previously discussed complexity results (**1.2.1, 1.2.2**) which revealed that the control group were producing fewer clauses and fewer words per AS-unit compared with the treatment group.

Table 17: percentage of error-free clauses per test

	groups	N	*Mean*	*SD*	*t*	*df*	95% Confidence Interval of the Difference *Lower*	*Upper*
Written pre-test/accuracy measure 1	planning	10	57.613	14.90	.284	18	12.854	16.870
	control	10	55.605	16.686			12.868	16.884
Written immediate post-test 1/ accuracy measure 1	planning	10	66.022	14.245	1.09	18	7.92	25.243
	control	10	57.363	20.498			8.071	25.389
Written delayed test 1/ accuracy measure 1	planning	10	74.097	13.310	1.45	18	5.670	31.296
	control	10	61.284	24.431			6.068	31.694
Written immediate post-test 2/ accuracy measure 1	planning	10	47.338	17.294	1.26	18	7.548	30.560
	control	10	35.832	22.879			7.650	30.662
Written delayed test 2/accuracy measure 1	planning	10	66.978	8.6398	2.01	12.39	.562	27.800
	control	10	53.359	19.518			1.035	28.273

4.1.2.8 <u>Accuracy measure 2: percentage of errors per test</u>

Results for the pre-test indicated that the percentage of errors the learners produced in both groups, including grammatical, lexical, and spelling errors, were almost identical. This indicated that all the learners were equal in terms of the accuracy of their L2 production prior to the beginning of the study. Having completed the text chat sessions, the results revealed that the treatment group produced greater immediate and delayed gains regarding this measure compared with the control group (see **table 18**). To put it simply, the former tended to make noticeably fewer errors than the latter, yet none of these results between the groups were found to be statistically significant. This meant that there appeared to be a positive effect of the feedback that treatment group learners received during the online sessions on their subsequent L2 production during the written tests. Longer and more frequent online sessions might have resulted in noticeable and significant progress in the accuracy of these learners' production. Additionally, the NSs/STs were instructed to try and address all the errors generated by the learners during the online sessions. This in turn might have somehow affected results across the tests and overburdened some learners with too many linguistic features to focus on while concurrently trying to attend to other aspects of their language (i.e., meaning, vocabulary, spelling). Hence, more structured and focused feedback, addressing a narrower range of particular linguistic forms at any one time, might have directed the learners' attentional resources to focus on different features of their L2 production more effectively.

Table 18: percentage of errors per test

	groups	N	*Mean*	*SD*	*t*	*df*	*Lower*	*Upper*
							95% Confidence Interval of the Difference
Written pre-test/accuracy measure 2	planning	10	10.276	3.7075	.044	18	4.45	4.26
	control	10	10.367	5.4146			4.49	4.31
Written immediate posttest 1/ accuracy measure 2	planning	10	9.3800	4.8711	.760	18	6.72	3.14
	control	10	11.167	5.6116			6.73	3.15
Written delayed test 1/ accuracy measure 2	planning	10	6.2170	3.1117	1.28	18	6.43	1.55
	control	10	8.660	5.1443			6.49	1.61
Written immediate posttest 2/ accuracy measure 2	planning	10	14.290	5.2529	.826	18	8.22	3.58
	control	10	16.612	7.1676			8.26	3.62
Written delayed test 2/ accuracy measure 2	planning	10	8.9510	3.9918	1.25	18	8.81	2.21
	control	10	12.247	7.2776			8.92	2.33

4.1.2.9 Fluency measure 1: percentage of words produced per minute

Overall, SLs in both groups produced more words as compared with the pre-test. This was expected given the fact that all the learners in this study had the chance to practise their L2 with more proficient L2 interlocutors during the online sessions. The treatment group, however, showed greater progress related to this measure as compared with the control group, and this difference between the groups was significant in delayed test 1 {t (18) =2.32, p=.03} (see **graph 5**). As stated earlier, the non-significant results for post-test 1 could be due to learners frequently editing and revising their written texts rather than lack of L2 fluency. Another significant difference between the groups was found for immediate post-test 2 {t (18) =2.6, p=.01}. This could be due to the explicit treatment delivered to the treatment group (i.e.,

NSs/STs were trying to help the learners be more fluent by providing different L2 chunks and structures), in addition to the direct influence of the voice chat mode.

Nevertheless, it is worth noting that the explicit instructions and feedback the treatment group received during the online sessions could have negatively affected the number of words they produced given the time they spent on the task. In other words, as the learners were trying to focus on different aspects of their production, they might have prioritized the quality of their production (in terms of grammar, structure, word choice, and number of errors) and paid less attention to the amount of words generated per each test.

Graph 5: percentage of words produced per minute
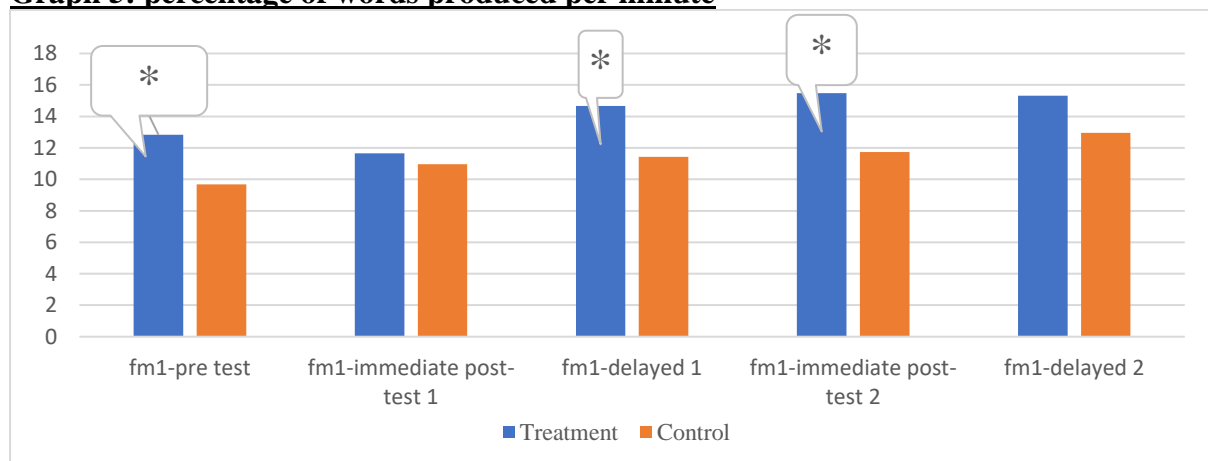
Note. **fm1= fluency measure 1, *= *p*≤ .05, **= *p*≤ .01**

4.1.3 Discussion of the SLs' written production results during the tests

CAF results in the current study show that the treatment presented to the experimental group during the online sessions (i.e., the planning time, the guided instructions, the explicit error-correction, and the manipulation of task type) did have a positive effect in developing learners' subsequent written performance throughout the immediate and delayed tests. Starting with *Syntactic complexity*:

- The treatment group achieved significant progress in the ratio of clauses per AS-unit as well as the length of AS-units the learners generated per test. This is in line with the findings of previous studies that examined the effects of pre-task planning on learners' written production, and which led to learners using significantly more complex

language (e.g., Kawauchi, 2005; Ellis & Yuan, 2004; Wendel, 1997; Foster & Skehan, 1996). Adams et al. (2015), however, who did not find any influence of task complexity on the ratio of clauses per AS-unit when analysing learners' writing, contend that their result cannot be a basis for generalization due to the short treatment duration; only one session for each group where the participants spent 45 minutes working on the interactive problem-solving task. Therefore, they concluded that longitudinal analysis of learners' interaction might lead to different results which, in fact, was the case in the current study.

▪ *Syntactic complexity* in this study was also operationalised in terms of learners' *production of complex grammatical forms*. Although the quantitative analysis of learners' written narratives during the tests showed that the experimental group outperformed the control group and produced more complex grammatical structures (such as relative clauses), none of the results between the groups were significant regarding this measure. The higher mean scores achieved by the treatment group lends general support to Mochizuki & Ortega's (2008) argument. They claim that when providing learners with external guidance to consider the integration of particular forms that might be essential, or at least useful, for task completion, learners' attention will be automatically oriented towards these forms. To the best of my knowledge, no previous studies have been conducted to gauge learners' production of relative clauses during written narrative tasks, and hence comparisons cannot be drawn.

▪ With regard to *syntactic variety*, the treatment group also showed greater progress and produced a greater variety of verb forms during the tests as compared with the control group, yet as stated earlier, the only significant result between the groups was found during delayed test 1. This is very similar to the findings of Ellis & Yuan (2004) who compared the effects of different planning conditions (pre-task planning, online

planning, and no planning) on the number of different grammatical verb forms learners used in their written narratives. They found that pre-task planning led to a marked and significant increase in the syntactic variety of learners' written production compared with the other two groups.

▪ Measures of ***Lexical variety*** (measured via means of MTLD) and ***lexical appropriacy*** were also employed to assess the complexity of learners' written narratives in this study. The experimental group also showed greater gains than the control group regarding the variety of the lexical items they produced throughout the tests; however, significant results were only found in immediate post-test 1 and delayed test 1. The treatment group produced more inappropriate lexical items compared to the control group and a significant difference between the groups was found during immediate post-test 2 and delayed test 2.

Previous studies also reported that the pre-task planning condition resulted in greater lexical variety compared with the no planning condition (g., Adams et al., 2015; Kawauchi, 2005). Similarly, not all the results from these studies were shown to be statistically significant. On the contrary, Ellis & Yuan (2004) found that pre-task planning had no effect on lexical variety. It is worth noting that Ellis & Yuan's findings were based on one treatment session and that they adopted Mean Segmental Type-Token Ratio (MSTTR) as a tool to assess lexical variety. Hence, the significant results in the current study and my contradictory findings when compared to Ellis & Yuan could be either due to the relatively longitudinal and explicit treatment received by the experimental group and/or the different tools used to measure the lexical variety of learners' narratives. Unfortunately, a full comparison with the findings of previous research concerning lexical variety was not possible given the different tools used to assess lexical variety, including TTR, number of word types, and Giraud index, to

mention but a few. Kawauchi (2005), however, found a significant interaction between planning conditions and learners' proficiency level; that is, more advanced learners produced a greater variety of lexical items when offered time to plan in advance. Hence, perhaps recruiting learners with a higher proficiency level could have led to more dramatic gains and significant results in the current study. Nonetheless, overall and for most of the complexity measures, the treatment group succeeded in producing more complex and syntactically/lexically varied language.

▪ General *accuracy measures* were used in this study; these involved *percentage of error-free clauses*; that is, the proportion of clauses that do not contain any type of errors) and *percentage of errors per test*. All morphological, syntactical, as well as lexical choice errors were counted. The treatment group had higher mean scores on the first measure (i.e., generated more error-free clauses) and produced fewer errors on the second measure compared with the control group. Independent samples t-tests, however, failed to show the differences between the groups to be statistically significant. These findings are in line with those of Asgarikia (2014). There were other studies, however, that found no effect of the planning time, provided prior to task's delivery, on the accuracy of the learners' writing performances when compared with no planning conditions (e.g., Seyyedi et al., 2013; Ziegler, 2018). These studies, therefore, suggest that when learners are given time to plan their task performance, they prioritise meaning over form (Sangarun, 2005).

▪ As previously mentioned, the *fluency* of SLs' written narratives across the tests was primarily assessed by calculating the number of words produced divided by the total number of minutes each SL used for task completion; this is typically referred to as 'production rate' (Skehan & Foster, 1999). The experimental group also outperformed the control group on this measure and significant results were detected in favour of the

former during delayed test 1 and immediate post-test 2. Again, this is consistent with the findings of previous studies (e.g., Allaw & McDonough, 2019; Ellis & Yuan, 2004; Seyyedi et al., 2013).

Summary of the CAF results in the written tests

To sum up, SLs achieved significant gains in the current study in terms of the complexity and fluency of their written production. However, no statistically significant results were captured with regard to the accuracy measures. In accordance with the findings from previous research, it seems that the effect of strategic planning on both complexity and fluency is greater than that on accuracy: learners lean towards drawing a conceptual plan of what they would like to say rather than being concerned about framing the linguistic details of their utterances, which in turn, seem hard to carry over into the actual task performance (Ellis, 2005; Seyyedi et al., 2013; Skehan, 1998; Yuan & Ellis, 2003). Thus, complexity and fluency are often favoured at the expense of accuracy due to trade-off effects (Foster & Skehan, 1996). The fact that learners' attentional resources are limited might have hindered their ability to simultaneously attend to all CAF components in the current study. The following sections will present and discuss the result of learners' oral production throughout the test.

4.1.4 Results of SLs' oral production during the pre-test

Independent samples t-tests were also conducted to assess learners' oral proficiency before the online sessions commenced and to uncover any significant differences between the treatment group and the control group in terms of the complexity, accuracy, and fluency of their L2 oral production (see **table 19** below). To this end, the same complexity (syntactic complexity, syntactic variety, lexical variety, and lexical appropriacy) and accuracy measures (percentage of error-free clauses, as well as percentage of errors per test) discussed earlier were employed.

Fluency measures, on the other hand, involved number of silent pauses, number of filled pauses, pruned speech rate, and repair measures.

Both groups were considered equal in terms of the complexity, accuracy, and fluency of their oral production during the spoken pre-test (see **graph 6 & 7**); although there were some differences between the groups in almost all the measures, none of these were found to be significant. One exception, however, was related to the duration of silent pauses made by the learners in both groups during the pre-test; the control group had a higher mean score (i.e., produced more silent pauses) than the treatment group. This difference between the groups was statistically significant (t (18) = 2.36, *p=*.02). While on the one hand this could imply that there was a gap between the groups in terms of the number of silent pauses they produced before the beginning of the online sessions, it could also be a natural outcome of other factors (e.g., individual factors). SLs rarely have a chance to practise their L2 in their contexts; hence some of them might have been more nervous than others during the test, and therefore may have decided to speak at a slower pace in order to have time to think of what to say next. Consequently, this might have resulted in a greater number of silent pauses on the part of the control group.

Table 19: descriptive statistics for CAF measures in the spoken pre test

CAF measures	Groups N= 10	Pre-test			
		M	*SD*	*Min*	*Max*
Ratio of clauses per AS-unit	TG	1.955	.642	1.21	3.20
	CG	1.893	.21		
Average length of AS-unit	TG	13.15	5.07	8.50	23.8
	CG	12.32	1.98		
Complex grammatical structures	TG	0.60	.699	0	3
	CG	0.70	.949		
Syntactic variety	TG	3.60	.966	2	5
	CG	3.50	.707		
Lexical variety	TG	41.45	14.39	26.61	73.28
	CG	36.77	8.84		
Lexical appropriacy	TG	0.10	.316	0	1
	CG	0.20	.422		
Error-free clauses	TG	60.26	13.14	26.66	82.35
	CG	51.57	15.53		
Number of errors per test	TG	9.19	3.34	3.35	21.90
	CG	12.91	6.31		
Duration of silent pauses	TG	.360	.085	.18	.51
	CG	.442*	.068		
Duration of filled pauses	TG	.169	.138	.0	.40
	CG	.159	.118		
Pruned speech rate	TG	125.19	20.0	71.0	155.0
	CG	112.92	26.21		
Mean duration of repairs	TG	.046	.033	.0	.10
	CG	.022	.028		

Overall then, all the SLs were considered homogenous in terms of their oral L2 proficiency at the beginning of this study, and thus, it would be easier to determine whether or not the guided planning and feedback received by the treatment group was more effective with regard to their L2 oral development.

Graph 6: Complexity and Accuracy results per group during the spoken pre-test
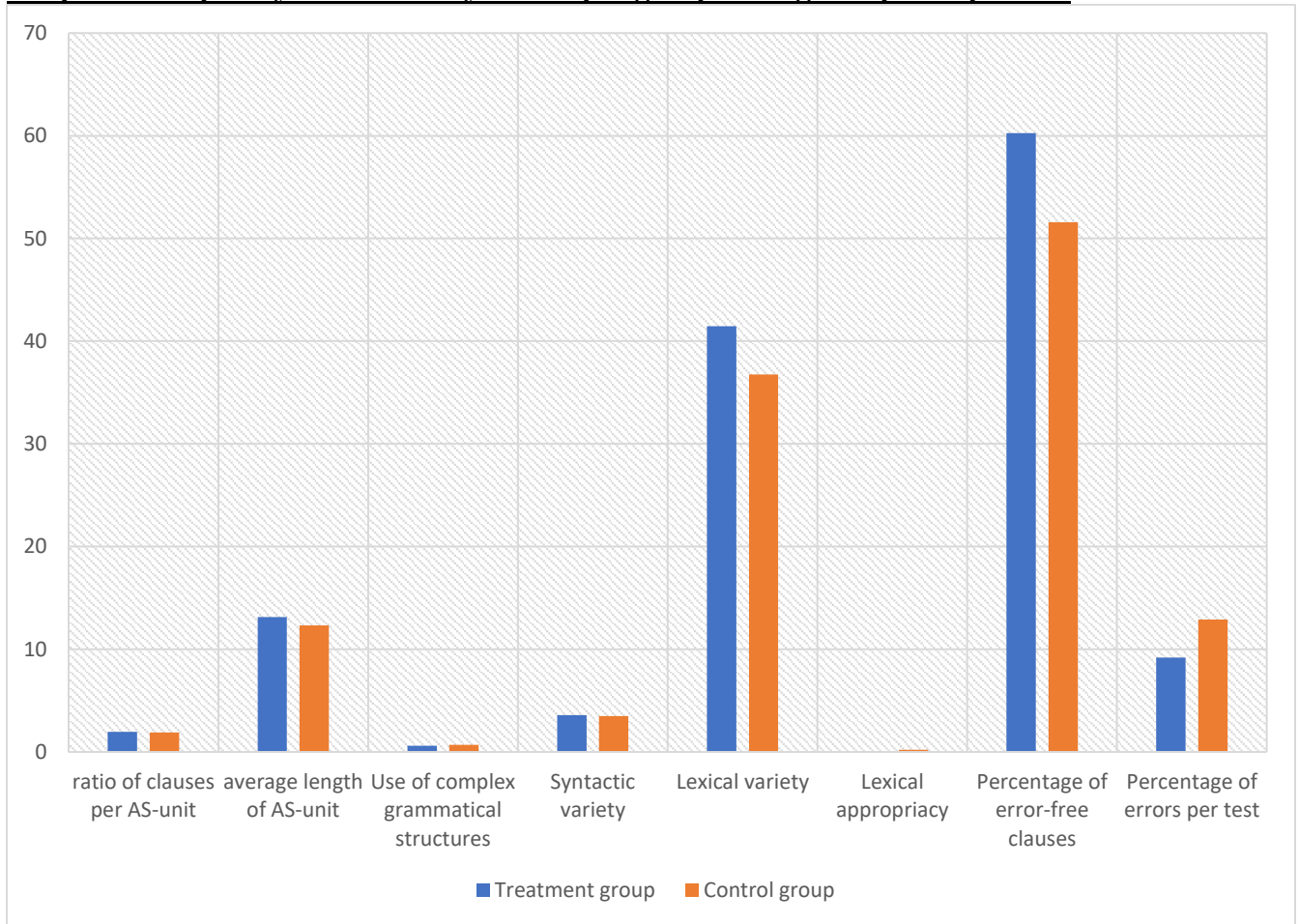
Graph 7: Fluency results per group during the spoken pre-test
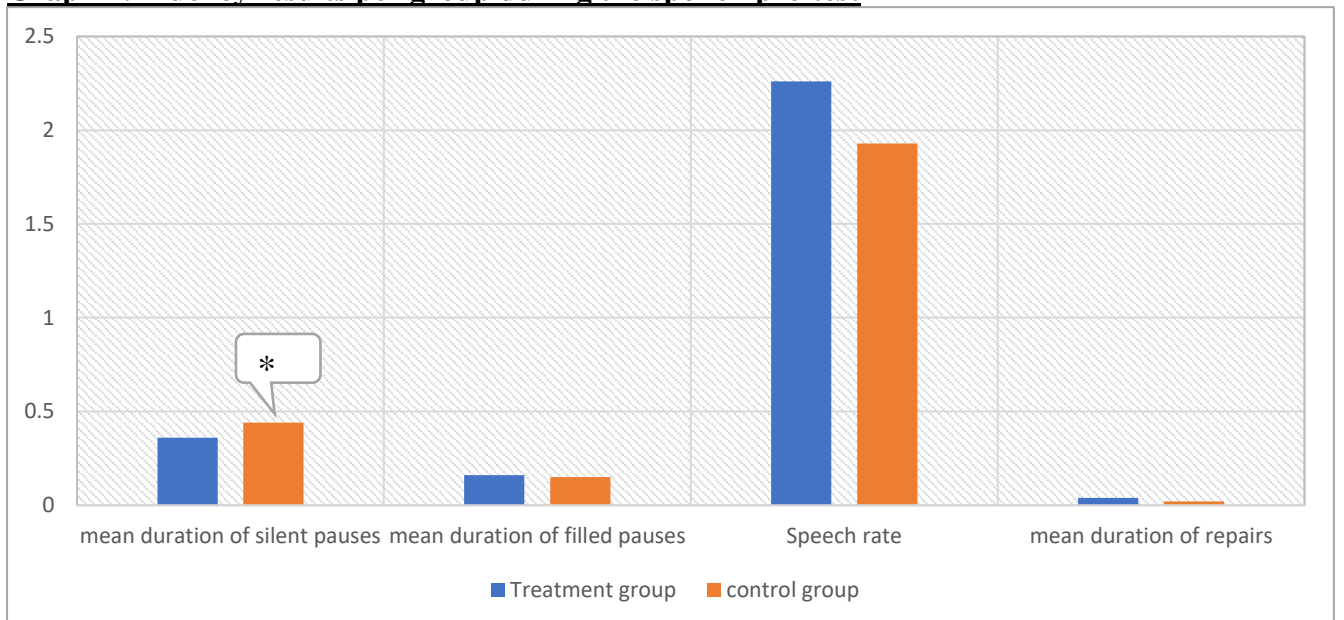
4.1.5 Results of SLs' oral production during the immediate and delayed post-tests:

Statistical tests, independent samples t-tests in particular, were carried out for each of the CAF measures to compare the linguistic performance of both groups during the immediate and delayed spoken post-tests. As the group statistics in **Tables (20 & 21), (22 & 23), (24 & 25)** show, overall and for almost all the CAF measures, the treatment group had better mean scores than the control group.

Table 20[21]: Descriptive statistics for Complexity measures: Treatment VS Control Group

Complexity measures	Group	Immediate post-test 1		Delayed test 1		Immediate post-test 2		Delayed test 2	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Ratio of clauses per AS-unit	TG	2.22	.38	2.78	.55	2.65	.57	2.81	.40
	CG	1.66	.27	1.87	.27	1.88	.28	1.75	.33
Average sentence length	TG	14.76	3.69	18.95	3.75	18.74	3.59	17.54	2.64
	CG	10.65	1.26	13.24	2.09	12.72	1.75	10.60	1.48
Complex grammatical structures	TG	.40	.516	.50	.707	.40	.699	.90	1.19
	CG	.40	.516	.20	.422	.40	.516	.40	.516
Syntactic variety	TG	3.60	1.26	3.40	.699	2.70	1.05	4.20	.919
	CG	3.30	1.16	3.40	.966	3.20	.632	3.70	1.16
Lexical variety	TG	39.9	9.06	39.35	11.25	43.75	12.48	38.41	4.68
	CG	35.8	10.05	30.13	6.49	41.02	12.16	38.47	11.52
Lexical appropriacy	TG	.70	.949	.20	.422	.10	.316	.30	.483
	CG	.50	.972	.40	.966	.20	.422	.40	.699

---

[21] **TG=treatment group, CG=control group, M=mean, SD=standard deviation, N=number of participants per group**

Table 21: Descriptive statistics for Complexity measures: Treatment VS Control Group

	Post-test 1		Delayed 1		Post-test 2		Delayed 2	
	min	*max*	*min*	*max*	*Min*	*max*	*min*	*max*
Ratio of clauses per AS-unit	1.14	2.83	1.55	4.00	1.50	3.66	1.36	3.50
Average length of AS-unit	8.40	22.17	10.2	27.3	10.4	24.2	8.18	23.2
Complex grammatical structures	0	1	0	2	0	2	0	3
Syntactic variety	2	6	1	4	1	4	2	6
Lexical variety	20.7	60.71	21.7	63.6	26.0	67.9	20.6	55.7
Lexical appropriacy	0	3	0	3	0	1	0	2

Table 22: Descriptive statistics for Accuracy measures: Treatment VS Control Group

Accuracy measures		Immediate post-test 1		Delayed test 1		Immediate post-test 2		Delayed test 2	
	Group N=10	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Error-free clauses	TG	67.81	16.34	58.30	19.36	59.64	19.59	70.10	11.62
	CG	45.33	21.41	44.76	15.97	46.55	22.45	49.21	14.59
Number of errors per test	TG	6.71	3.28	10.79	6.19	8.25	5.23	7.37	3.09
	CG	12.09	6.01	13.86	5.55	13.03	6.48	16.54	7.28

Table 23: Descriptive statistics for Accuracy measures: Treatment VS Control Group

	Immediate post-test 1		Delayed 1		Immediate post-test 2		Delayed 2	
	min	*max*	*min*	*max*	*Min*	*max*	*min*	*max*
Error-free clauses	21.0	92.30	26.6	84.2	12.5	88.2	25.0	81.4
Number of errors per test	1.42	22.01	2.63	21.5	1.85	25.2	4.63	31.7

Table 24: Descriptive statistics for Fluency measures: Treatment VS Control Group

Fluency measures	Group N=10	Immediate post-test1		Delayed test1		Immediate post-test2		Delayed test2	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
duration of silent pauses per task	TG	.39	.044	.34	.103	.32	.135	.30	.047
	CG	.41	.045	.43	.076	.36	.078	.35	.047
duration of filled pauses per task	TG	.17	.133	.18	.152	.16	.142	.12	.10
	CG	.13	.112	.14	.114	.12	.077	.11	.10
Pruned Speech rate	TG	125.4	21.1	121.9	22.61	136.5	21.38	155.0	25.99
	CG	114.58	28.6	112.6	23.20	135.8	28.2	131.1	19.82
Duration of repairs	TG	.033	.031	.030	.034	.043	.03	.024	.029
	CG	.029	.040	.001	.003	.0	.0	.0	.009

Table 25: Descriptive statistics for Fluency measures: Treatment VS Control Group

	Post-test 1		Delayed 1		Post-test 2		Delayed 2	
	min	*max*	*min*	*max*	*Min*	*max*	*min*	*max*
Mean duration of silent pauses	.33	.53	.21	.56	.08	.48	.20	.43
Mean duration of filled pauses	.00	.41	.01	.55	.00	.46	.00	.33
Pruned speech rate	77.0	179.9	84.0	167	92.0	176	87.0	198
Mean duration of repairs	.00	.10	.00	.10	.00	.10	.00	.10

To put it simply, the treatment group made greater progress according to the test results (i.e., immediate and delayed) in terms of their oral L2 proficiency as compared with the control group. This reveals that manipulating planning time and explicitness of instruction (i.e., instructions and feedback) had an effect, though not a very strong effect for all the measures. The following paragraphs will present and elaborate on the findings associated with the analysis of each of the CAF measures across the immediate and delayed tests.

4.1.5.1 <u>Complexity measure 1: ratio of clauses per AS-unit</u>

T-tests results revealed a difference between the groups in terms of the number of clauses they generated per AS-unit. To put it another way, the treatment group produced a greater number of clauses per AS-unit than the control group throughout the tests, and this difference between the groups was found to be statistically significant: immediate posttest 1(t (18) =3.72, p=.002), delayed test 1 (t (18) =4.64, p=.000), immediate posttest 2 (t (13.19) =3.83, p=.002), and delayed test 2 (t (18) =6.37, p=.000) (see **graph 8**). Hence, there was strong evidence to suggest that the treatment received by the guided planning group, including planning time as well as explicit feedback/instructions, had a significant effect on the number of clauses they produced per AS-unit and their ability to generate more syntactically complex sentences.

Graph 8: ratio of clauses per AS-units per oral test
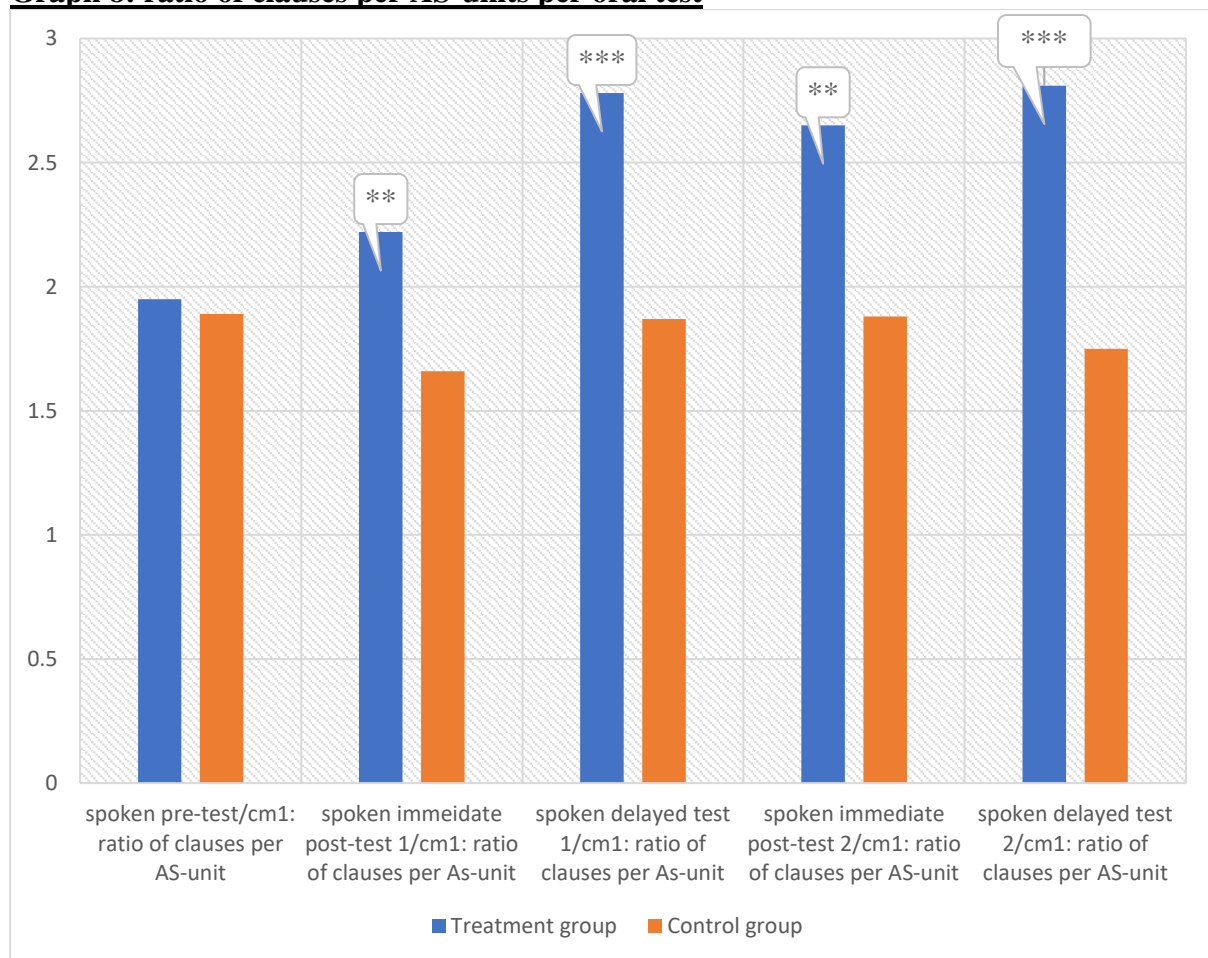
4.1.5.2 Complexity measure 2: average length of AS unit

In line with the findings concerning the previous measure, similar results were uncovered for the second complexity measure. That is to say, the treatment group were producing longer AS-units than the control group during immediate post-test 1. The effects of the treatment received by the guided planning group and the explicit feedback delivered by their more proficient L2 interlocutors resulted in the learners' tendency to produce more complex and extended rather than simple sentences across all the post- and delayed tests as compared with the control group. Again, there was very strong evidence to conclude that this difference between the groups was statistically significant: immediate posttest 1 (t (11.08) =3.32, p=.007), delayed test 1 (t (18) = 4.20, p=.001), immediate posttest 2 (t (13.04) = 4.76, p=.000), and delayed test 2 (t (18) =7.24, p=.000) (see **graph 9**).

Graph 9: Average length of AS-unit per oral test

150
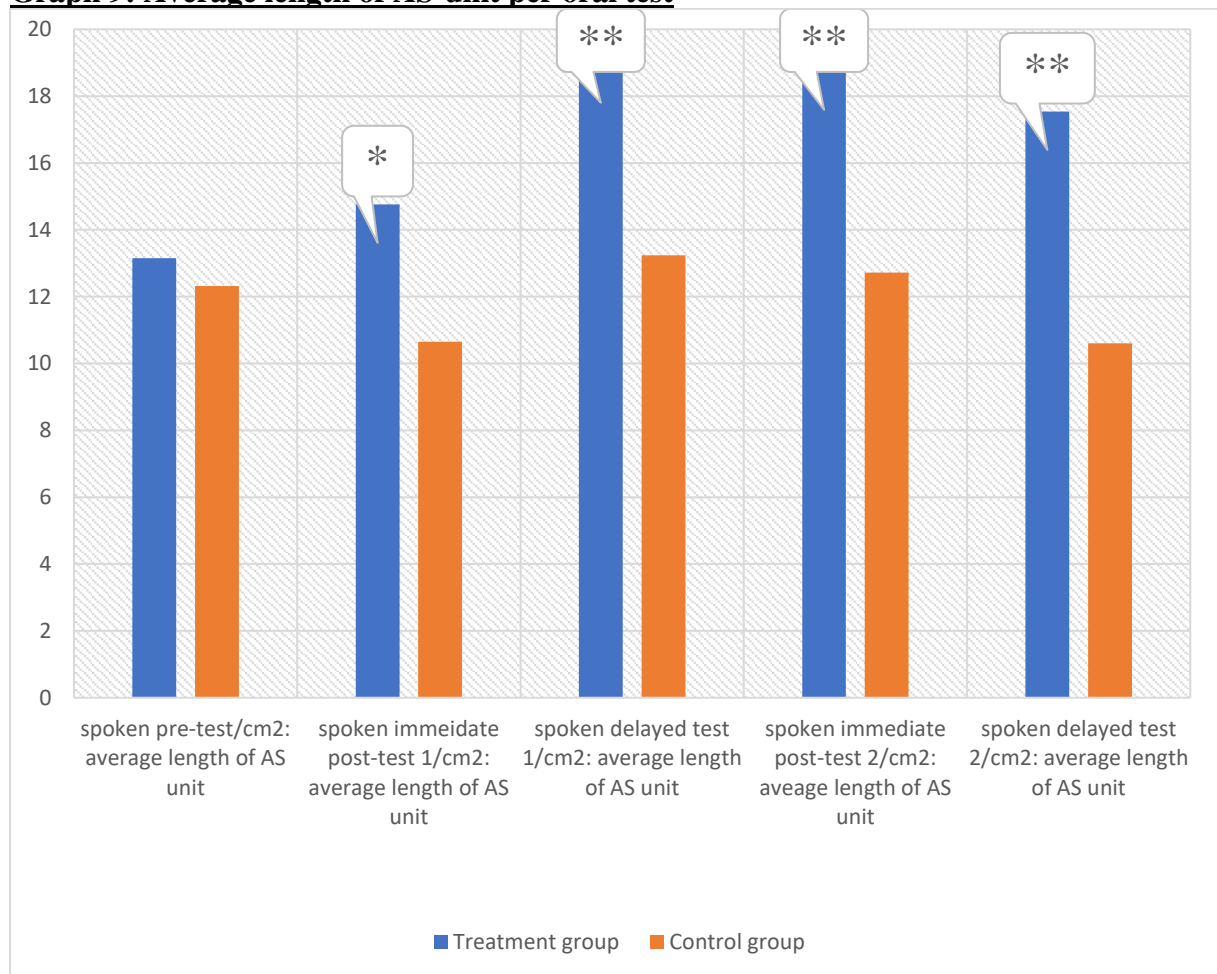
4.1.5.3 Complexity measure 3: learners' production of complex grammatical structures

Learners in the treatment group generated either a similar or greater number of relative clauses as compared to the control group in all the post- and delayed tests (see **table 26**; only error-free relative clauses were calculated). Unlike the findings of the written tests (immediate post-test 2 in particular), no statistically significant differences between the groups were found regarding this measure during the oral tests.

Table 26: number of complex grammatical structures per oral test

	groups	N	*M*	*SD*	*t*	*df*	*95% Confidence Interval of the Difference*	
							Lower	*Upper*
Spoken pre-test/ complexity measure 3; number of complex grammatical structures	planning	10	.60	.699	.268	18	.883	.683
	control	10	.70	.949			.888	.688
Spoken immediate post-test 1/ complexity measure 3; number of complex grammatical structures	planning	10	.40	.516	.000	18	.485	.485
	control	10	.40	.516			.485	.485
Spoken delayed test 1/ complexity measure 3; number of complex grammatical structures	planning	10	.50	.707	1.152	14.682	.247	.847
	control	10	.20	.422			.256	.856
Spoken immediate post-test 2/ complexity measure 3; number of complex grammatical structures	planning	10	.40	.699	.000	18	.577	.577
	control	10	.40	.516			.581	.581
Spoken delayed test 2/ complexity measure 3; number of complex grammatical structures	planning	10	.90	1.197	1.21	18	.366	1.36
	control	10	.40	.516			.396	1.39

The complexity of these forms as well as the mode of delivery (i.e., speaking rather than writing) could explain the non-significant results, especially for the treatment group, which was receiving explicit instructions and feedback during the online sessions related to their production of relative clauses. To put it another way, while speaking the learners might have been unable to integrate complex structures (such as relative clauses) into their output because they had not fully mastered or fully acquired these and had only a limited time available to think of the syntactic structure of their L2 production during the tests. Consequently, they might have preferred to be on the safe side and only produce the forms they were confident about. Additionally, although NSs/STs tried to be more explicit in their feedback during the voice chat sessions, they might have chosen to avoid frequently commenting on learners' production of relative clauses in order not to interrupt the flow of the interaction. The short duration of the online sessions could be another reason for the non-significant results; complex grammatical structures might require longer sessions and more focused and explicit treatment to acquire.

4.1.5.4 <u>Complexity measure 4: syntactic variety</u>

As mentioned earlier, this was measured by calculating the overall number of different verb forms (i.e., tense, voice, and modality) the learners produced per task. Independent samples t-tests showed that throughout almost all the tests, the treatment group had higher mean scores (see **table 27**), and hence produced a greater variety of verb forms as compared with the control group. However, no significant difference was detected between the groups regarding this measure.

Table 27: number of different verb forms per oral test

	groups	N	*M*	*SD*	*t*	*df*	95% Confidence Interval of the Difference Lower	Upper
Spoken pre-test/ complexity measure 4; number of different verb forms	planning	10	3.60	.966	.264	18	.695	.895
	control	10	3.50	.707			.701	.901
Spoken immediate post-test 1/ complexity measure 4; number of different verb forms	planning	10	3.60	1.265	.553	18	.840	1.44
	control	10	3.30	1.160			.841	1.44
Spoken delayed test 1/ complexity measure 4; number of different verb forms	planning	10	3.40	.699	.000	18	.792	.792
	control	10	3.40	.966			.798	.798
Spoken immediate post-test 2/ complexity measure 4; number of different verb forms	planning	10	2.70	1.059	1.282	14.693	1.32	.320
	control	10	3.20	.632			1.33	.333
Spoken delayed test 2/ complexity measure 4; number of different verb forms	planning	10	4.20	.919	1.069	18	.483	1.48
	control	10	3.70	1.160			.487	1.48

The lack of significant results between the groups could be traced back to different factors. First, the mode of task delivery, i.e., the learners had only a very short time to view the pictures and deliver a verbal narration near-simultaneously. Thus, the time factor might have inhibited the learners' use of different and complex verb forms. The task type employed during the tests could be another reason: SLs worked on one task type (storytelling activities) throughout the tests and hence they might have tended to use simple tenses to describe what they were seeing in the pictures. Introducing different task types into the study design could have therefore resulted in a greater variety of syntactic structures. Individual differences amongst the learners

could have also led to the non-significant results; alternatively, learners' understanding of the content of each task possibly pushed them to employ particular forms, while on the other hand, made them intentionally avoid others. To put it simply, when the learner easily understood the sequence of the incidents, s/he may have been more comfortable using a wider range of various syntactic structures. On the contrary, when s/he struggled to understand the story unfolding in the pictures, s/he may have preferred to stick to simpler language and simpler syntactic structures.

4.1.5.5 <u>Complexity measure 5: lexical variety</u>

The measure of textual lexical diversity (MTLD) was employed to assess the lexical richness of learners' oral production. Thus, the types or the number of different words generated by the learner were divided by its tokens, i.e., the total number of words produced per test. The statistical tests revealed that the treatment group produced a greater variety of lexical items than the control group in almost all the tests (except for delayed test 2, where the two groups had approximately the same amount of lexical variation in their oral narrations). The only significant result between the groups, however, was found during delayed test 1 (t (18) $= 2.23$, $p= .03$) (see **graph 10**). One plausible reason for the non-significant results between the groups in the other tests could be that all the SLs had the chance to practise their language with more proficient L2 interlocutors. Accordingly, learners working in the control group could have also benefited from the online interaction with their L2 partners and learned new lexical items. Nevertheless, this does not necessarily mean that all these items were accurate and meaningful given the context of the task. This will be further discussed in the following section. Another explanation could be that the tasks' content varied throughout the study, i.e., most of the words generated per task were task-related, and therefore the treatment group were not able to find an opportunity to use all the words they learned from their partners during the tests.

Graph 10: lexical variety per oral test
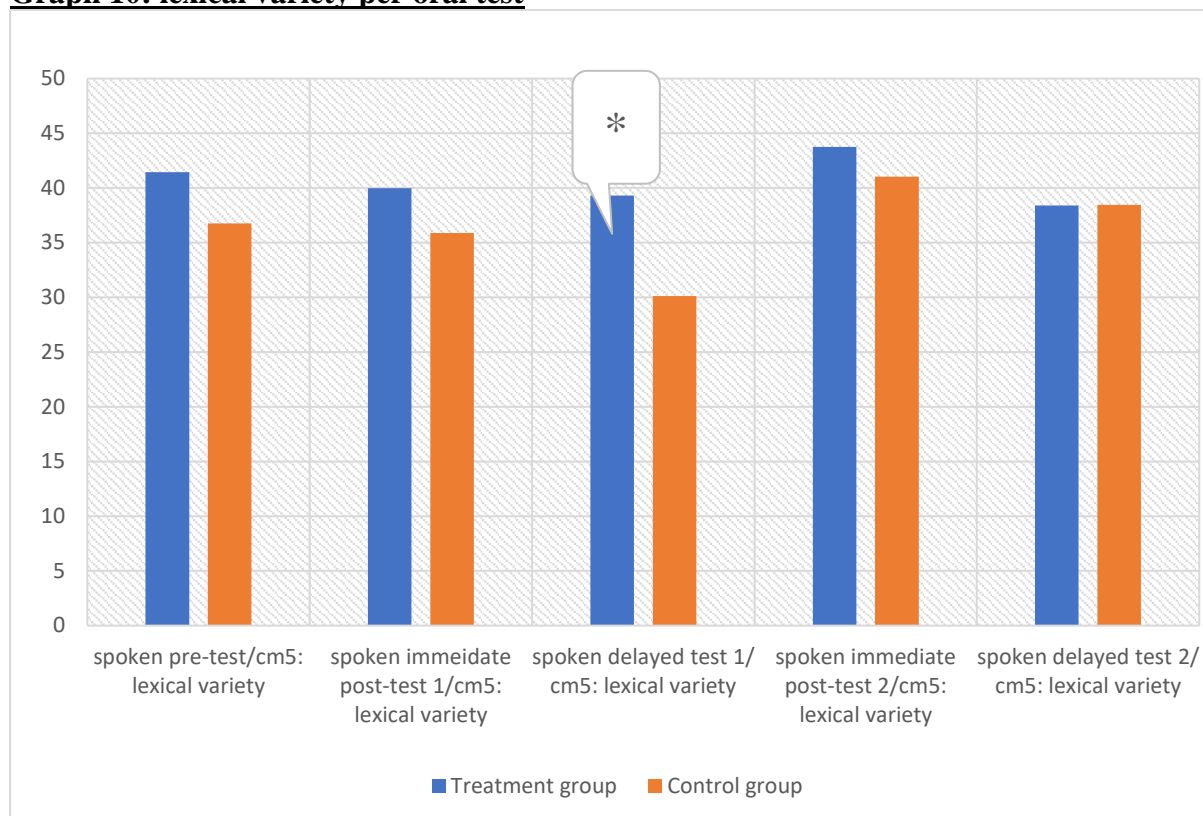
4.1.5.6 Complexity measure 6: lexical appropriacy

All the lexical items produced by the learners per test and which were found to be inappropriate in terms of meaning given the context of the narrative were counted. In agreement with the results of the previous measure, independent samples t-tests showed that the control group produced more inappropriate words and had higher mean scores (see **table 28**) compared with the treatment group throughout the tests, except for immediate post-test 1; yet none of these differences were statistically significant.

Table 28: lexical appropriacy per oral test

	groups	N	M	SD	t	df	95% Confidence Interval of the Difference Lower	Upper
Spoken pre-test/ complexity measure 6; lexical appropriacy	planning	10	.10	.316	.600	18	.450	.250
	control	10	.20	.422			.452	.252
Spoken immediate post-test 1/ complexity measure 6; lexical appropriacy	planning	10	.70	.949	.466	18	.702	1.10
	control	10	.50	.972			.702	1.10
Spoken delayed test 1/ complexity measure 6; lexical appropriacy	planning	10	.20	.422	.600	18	.900	.500
	control	10	.40	.966			.924	.524
Spoken immediate post-test 2/ complexity measure 6; lexical appropriacy	planning	10	.10	.316	.600	18	.450	.250
	control	10	.20	.422			.452	.252
Spoken delayed test 2/ complexity measure 6; lexical appropriacy	planning	10	.30	.483	.372	18	.665	.465
	control	10	.40	.699			.670	.470

Hence, although as shown earlier the learners in the control group made some progress following the voice chat online sessions and produced a slightly greater variety of lexical items during delayed test 2, it seems that not all these items employed by the control group were adequate for the given context. The reason behind the treatment group generating more inappropriate words than the control group during immediate post-test 1 could be because of the effects of the text-based chat sessions. The learners were seeing and noticing the variety of lexical items employed by their NS/ST partners on the screen; and therefore, they might have been very excited and keen on integrating these items into their L2 production, not paying attention to the fact that different tasks might require a different set of lexical items.

Summary of the results for oral complexity:

With regard to *syntactic complexity* (ratio of clauses per AS-unit and the average length of AS-unit),

- the findings revealed that the treatment group outperformed the control group throughout the immediate and delayed post-tests and that all the results were statistically significant. In line with the findings of previous research (e.g., Yuan & Ellis, 2003), this indicates that the pre-task planning as well as the feedback provided during the online sessions had a strong effect on the syntactic complexity of the treatment group's L2 production throughout the tests. However, Révész (2011) revealed that the overall syntactic complexity of learners' oral production in F2F interaction decreased with a more complex version of a task. Révész concluded that this result is in line with Robinson (2005), who predicts "lower structural complexity on more complex interactive tasks owing to the amplified amount of negotiation that tends to result when the cognitive demands of interactive task are increased" (p.176). Perhaps the fact that the treatment group were practising their language online with more proficient L2 interlocutors and receiving feedback on their language use positively affected learners' subsequent L2 performance throughout the tests. This could explain the non-significant findings revealed in previous SCMC research related to syntactic complexity (e.g., Abrams, 2003; Kost, 2004). In these two studies, the learners did not receive guided planning prior to the online interaction or benefit from the language and feedback delivered by more proficient L2 interlocutors throughout the sessions.

- Although the treatment group made some progress in terms of *the number of complex grammatical structures* (relative clauses) unlike previous research (e.g., Mochizuki & Ortega, 2008; Thompson, 2014), none of the results were statistically significant

between the groups. The discrepancies between the findings of the current study and those of previous research could be due to many factors. First, Mochizuki & Ortega's study involved a one-off experiment that examined the effect of pre-task planning at one point in time. Second, the discrepancies could be explained by the content of the narrative tasks: Thompson (2014), for example, used narrative tasks that had almost the same content or flow of ideas (a family in a shop trying to buy a pet, a car, shoes, etc.) and the task's complexity was manipulated by varying the number of relative clauses and variety of verb forms the learners needed to produce per task. Hence, the gains the learners achieved occurred from taking some time to rehearse and plan the targeted forms as well as practise using them throughout the study. However, in the case of the current research, SLs were given narrative tasks which resembled everyday situations, and which were entirely different in terms of content and flow of events. Therefore, the learners' attempts to understand the tasks and attend to different aspects of their production, given the limited time they were offered to complete the tasks (i.e., only one minute), could explain the non-significant results for some of the measures.

- Similarly, the treatment group produced a wider variety of ***different grammatical verb forms*** per test than the control group, yet no statistically significant difference was found between the groups. These results are in line with those of Yuan & Ellis (2003), who examined the effects of pre-task planning/online planning on CAF. They found that both planning groups outperformed the non-planning group in terms of the syntactic variety of their L2 production. Again, none of these comparisons were statistically significant.

- Vercellotti (2017), who examined the linguistic performance of individual learners during multiple topic-based speeches, found that speeches with higher ***lexical variety*** scores had longer AS-units. However, the fact that different types of tasks were

employed during the online sessions (including spot the difference, narrative, and decision-making tasks) could have made it difficult for the SLs to retrieve the lexical items they had been exposed to by their partners and integrate them into their subsequent L2 production throughout the test. Therefore, they might have preferred not to risk jeopardising the accuracy of their narrations by employing lexical items they were not very confident about using. This could explain the lack of steady progress and significant results between the groups regarding the lexical variety and *lexical appropriacy* of the learners' production across the immediate and delayed tests. This result also seems in accordance with previous CAF research that examined lexical variety in oral production (e.g., Yuan & Ellis, 2003).

4.1.5.7 Accuracy measure 1: percentage of error-free clauses

This was measured by dividing the number of error-free clauses by the total number of clauses produced per test, multiplied by 100:

(Number of error-free clauses produced per test)   ***100**
(Number of clauses produced per test)

Findings obtained from the statistical tests revealed that there was a difference between the groups: the treatment group produced more error-free clauses than the control group across all the immediate and delayed tests. This result is expected given the explicit instructions as well as the feedback SLs in the treatment group received during the online sessions to focus on their L2 production. The only statistically significant differences were found, however, in immediate post-test 1 t (18) = 2.63, p= .01) and in delayed test 2 (t (18) = 3.54, p= .002) (see **graph 11**). On the one hand, the lack of significant results between the groups in the other tests could possibly be traced back to various reasons: the treatment group trying to focus on different aspects of their production (i.e., form, meaning, pronunciation), the learners delivering the task orally which might have put extra pressure on their ability to focus on their errors, and/or the

159

short duration of the online sessions that probably did not provide learners sufficient time to make progress with regard to the number of error-free clauses they produced. The significant results associated with immediate post-test 1, on the other hand, could be due to the direct impact of the text-based chatting sessions where the learners had the chance to notice their errors on their screens as highlighted and corrected by their more proficient L2 interlocutors. Additionally, the significant progress achieved by the treatment group when compared with the control group in delayed test 2 was perhaps because the SLs had built up their L2 knowledge as the study was progressing and managed to gradually improve their ability to access different aspects of their L2 production.

Graph 11: percentage of error-free clauses per oral test
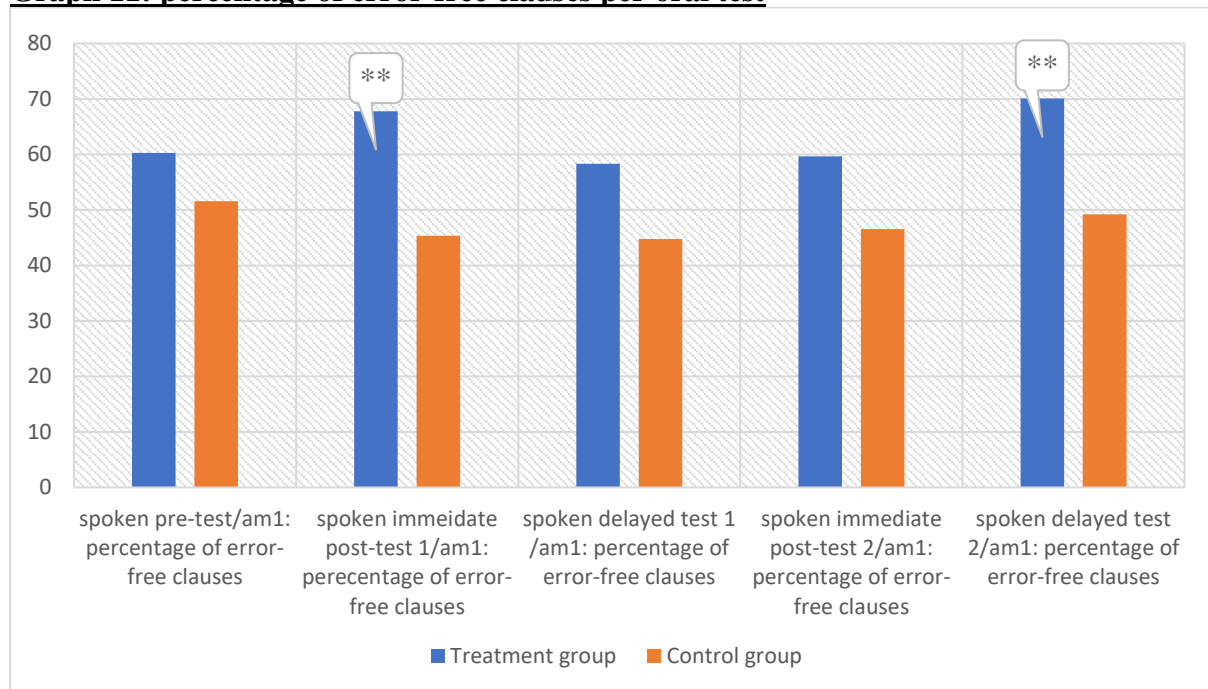
4.1.5.8 Accuracy measure 2: percentage of errors per test

This was calculated by dividing the number of errors, including grammatical and lexical errors, by the total number of words the learners produced per test, then multiplying by 100.

(Number of errors produced per test)   *100
(Number of words produced per test)

Overall, the treatment group produced fewer errors given the total number of words they generated per test as compared with the control group. Again, the only significant differences were detected during immediate post-test 1 (t (18) = 2.48, p=.02) and delayed test 2 (t (18) = 3.66, p=.002) (see **graph 12**). This was found to be compatible with the findings of the previous accuracy measure (the higher the percentage of error-free clauses, the lower was the percentage of errors given the total number of words produced per test), and hence the results could be justified accordingly. Thus, there is strong evidence to suggest that the treatment group was benefiting from the explicit treatment they received during the online sessions and achieving greater progress than the control group in terms of the accuracy of their oral L2 production.

Graph 12: percentage of errors per oral test
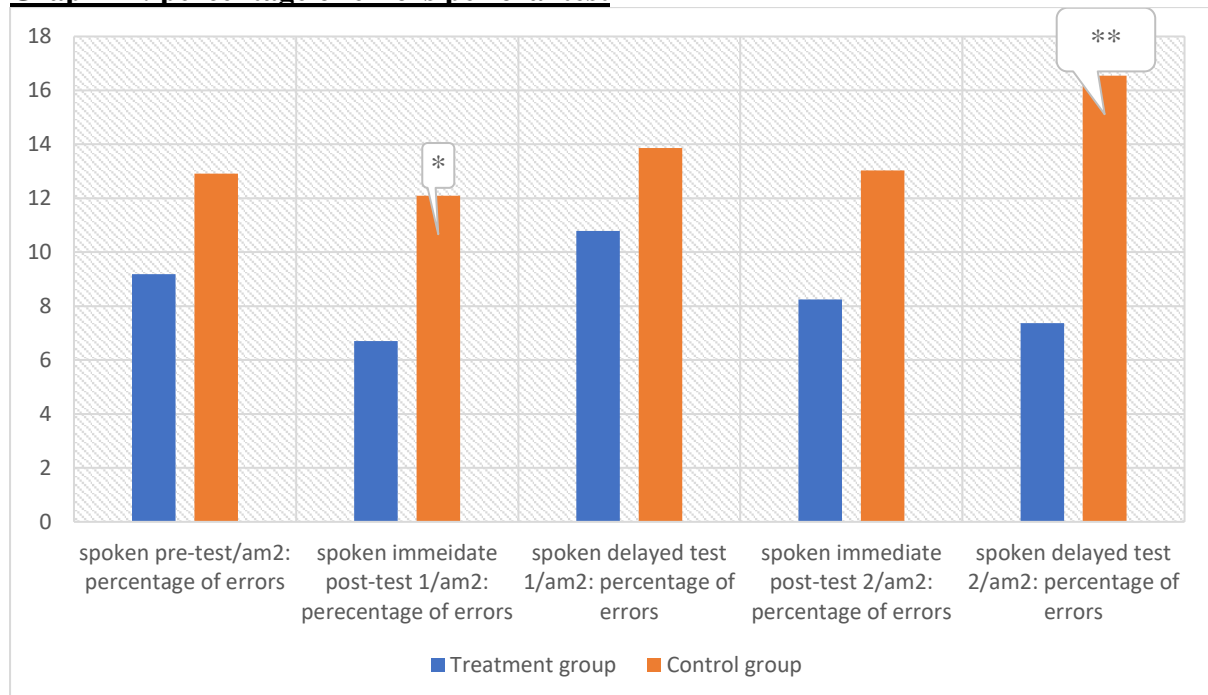
Summary of the results for oral accuracy

To sum up, general measures were utilized in this study to gauge learners' progress, if any, and determine how accurate their L2 production was throughout the tests. The findings revealed that both the guided planning as well as the explicit feedback the treatment group received

before and during the online sessions were contributing factors to the significant gains they achieved in accuracy during immediate post-test 1 and delayed test 2. Given the learners' proficiency level, it appears that L2 practice opportunities were not as effective for developing accuracy as continual guided planning and explicit feedback. To put it another way, continual guidance and scaffolding the learners as they progressed with the online sessions could have amplified learners' developmental gains in terms of accuracy. More significant results were reported in previous research (e.g., Yuan & Ellis, 2003). The inconsistencies between the findings of the current study and those of Yuan & Ellis (2003) were possibly due to the conditions under which the learners in the latter carried on the tasks (i.e., planning time and five minutes to orally narrate the story). So, unlike this study (where learners only had one minute to complete the test), the participants in Yuan and Ellis's (2003) study had plenty of time to plan and think about their utterances. One strength of the quantitative findings of the current research, especially those related to the oral tests, is that they provide a naturalistic picture of how learners' L2 production developed over time. In other words, SLs were working under conditions that resembled real-life conversation as they had not been offered time to plan or think about what they would say during the tests; all the learners were given just one minute to achieve the task's goal.

4.1.5.9 <u>Fluency measure 1: duration of silent pauses</u>

Following Révész et al. (2016), this was measured by dividing the number of unfilled pauses (i.e., total silence) greater than 0.25 seconds by the amount of time the learners spent on the test (i.e., speaking time). This is one of the measures commonly used to assess breakdown of fluency (e.g., Tavokoli & Skehan, 2005; Bui & Skehan, 2018) and consequently to determine learners' disfluency by gauging how often they paused during the test.

Figure 10: A screenshot of the speech analysis of learners' oral tests
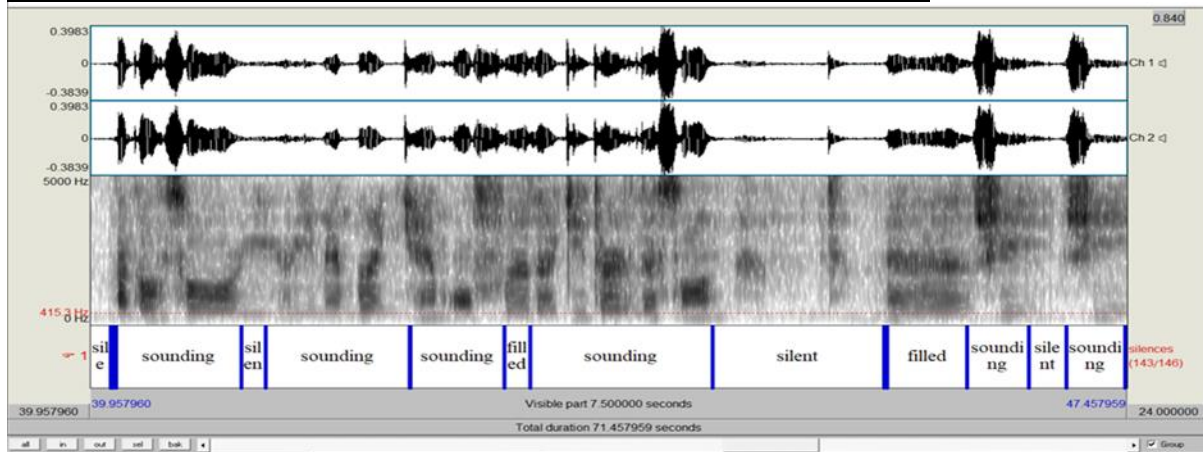
Figure 10 reproduces a screenshot taken from Praat (Boersma & Weenink, 2013), a free computer software package which was used to do speech analysis for the oral tests conducted by the learners in the current study. The application aids the transcription of speech samples by displaying waves and lines that resemble sounding or pausing segments. These waves and lines were manually labelled as sounding boundaries to refer to the words/syllables the learner produced or silent/filled boundaries to point to the type of pauses recognised in learners' oral production. Independent samples t-tests revealed a difference in the mean scores between the groups (see **table 29**); the treatment group had lower mean scores for silent pauses (producing relatively fewer pauses) than the control group. Statistically significant results, however, were only associated with delayed test 1 and delayed test 2, (t (18) = 2.18, $p=.04$) and (t (18) = 2.34, $p=.03$) respectively. The learners in the treatment group might have been influenced by the instructions/feedback they received during the online sessions, which in turn could have caused the non-significant results between the groups during the immediate tests. In other words, the experimental group were trying to take the feedback they received online from their partners on board and attend to different facets of their production when performing the immediate post-tests. That is, they needed more time to achieve the task goal and hence more pauses were uncovered as a result. Longer and more frequent treatment duration could have led to more statistically significant differences between the groups across all the tests.

Table 29: duration of silent pauses

	groups	N	*M*	*SD*	*t*	*df*	Lower	upper
							| 95% Confidence Interval of the Difference |	
Spoken pre-test/fm1: duration of silent pauses	planning	10	.3600	.0856	2.36*	18	.154	.009
	control	10	.4420	.0682			.155	.008
Spoken immediate post-test 1/fm1: duration of silent pauses	planning	10	.3930	.0449	1.18	18	.066	.018
	control	10	.4170	.0457			.066	.018
Spoken delayed test 1/ fm1: duration of silent pauses	planning	10	.3490	.1035	2.18*	18	.174	.003
	control	10	.4380	.0762			.174	.003
Spoken immediate post-test 2/fm1: duration of silent pauses	planning	10	.3220	.1355	.767	14.422	.142	.066
	control	10	.3600	.0784			.143	.067
Spoken delayed test 2/fm1: duration of silent pauses	planning	10	.3090	.0477	2.34*	18	.094	.005
	control	10	.3590	.0474			.094	.005

Note. **fm1 = fluency measure 1**

4.1.5.10 Fluency measure 2: mean duration of filled pauses

Another measure used in the present study to assess breakdown fluency was the mean duration of filled pauses. This was obtained by calculating the total number of filled pauses, i.e., pauses where the learners used 'Erm', 'Er', and so on for a duration greater than 0.25 seconds, and then dividing it by the speaking time the learners spent on the task. The statistical tests showed that the treatment group had higher mean scores (see **table 30**), i.e., produced more filled pauses, compared with the control group during the immediate and delayed tests. No statistically significant differences were found between the two groups regarding the number of filled pauses they produced per test, however. With regard to the high mean scores of filled pauses produced by the planning group, these pauses could be traced back to the impact of the

treatment the learners received during the online sessions. Hence, they were taking some time to think of the structure of their utterances while performing the tests. This seems in line with the complexity findings discussed in **sections 1.5.1 and 1.5.2**, which revealed that the treatment group produced syntactically more complex language than the control group throughout the oral tests.

The context of the learners could have also influenced the number of pauses they produced in the current study. To put it another way, at the time of the study they did not experience any external exposure to the language, or any chance to practise their L2 in their own context. Additionally, the frequency, the number, and the duration of each voice chat session the learners had with more proficient L2 users might not have been enough to result in significant progress on the number of pauses (either filled or unfilled) the learners made per test.

Table 30: mean duration of filled pauses

	groups	*N*	*M*	*SD*	*t*	*df*	*95% Confidence Interval of the Difference Lower*	*Upper*
Spoken pre-test/fm2: duration of filled pauses	planning	10	.1690	.13828	.174	18	.111	.131
	control	10	.1590	.11874			.111	.131
Spoken immediate post-test 1/fm2: duration of filled pauses	planning	10	.1740	.13327	.688	18	.078	.154
	control	10	.1360	.11296			.078	.154
Spoken delayed test 1/ fm2: duration of filled pauses	planning	10	.1810	.15249	.581	18	.091	.161
	control	10	.1460	.11404			.092	.162
Spoken immediate post-test 2/fm2: duration of filled pauses	planning	10	.1670	.14268	.836	18	.065	.151
	control	10	.1240	.07792			.067	.153
Spoken delayed test 2/fm2: duration of filled pauses	planning	10	.1240	.10501	.173	18	.089	.105
	control	10	.1160	.10157			.089	.105

4.1.5.11 Fluency measure 3: 'pruned speech rate'

Pruned speech rate refers to the average number of meaningful syllables produced per minute, and from which repetitions, false starts and other performance features (e.g., self-corrections, L1 use, incomprehensible language) have been excluded. According to Tavakoli & Skehan (2005), this measure "deals with the speed with which language is produced" (p. 254). Accordingly, this measure helped in measuring how fast learners produce L2 speech in the current study. Pruned speech rate was measured as follows:

(Total number of meaningful syllables)
(Total number of seconds)   ***60**

Independent samples t-tests showed that the treatment group was generating more syllables given the time they spent on the task, and consequently was more fluent when compared with the control group. Yet the only statistically significant difference found between the groups was for delayed test 2 (t (18) = 2.30, $p=.03$) (see **graph 13**). The increased rates for the treatment group were because of the additional preparation time/instructions they received prior to attempting the tasks in the online session. Learners' involvement with the task as well as their attempts to simultaneously attend to multiple L2 features (e.g., meaning, grammar) could have affected the learners' fluency and therefore might explain the lack of significant results throughout the other tests. As mentioned earlier, more frequent L2 practice with more proficient L2 users could have resulted in more significant differences.

Graph 13: pruned speech rate
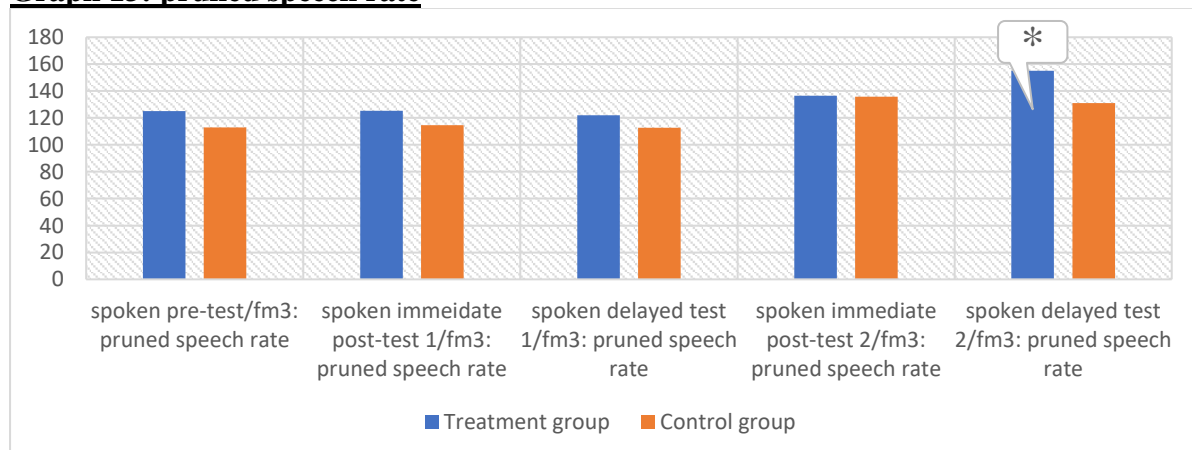
4.1.5.12 Fluency measure 4: mean duration of repairs

This was calculated via means of dividing the overall number of repairs the learners made per test (including reformulations, repetitions and so on) by the total time needed to complete the assigned task. As shown in **table 31** below, the treatment group had higher mean scores and hence made more repairs during the immediate and delayed tests compared with the control group.

Table 31: mean duration of repairs

	groups	N	*M*	*SD*	*t*	*df*	*Lower*	*Upper*
							95% Confidence Interval of the Difference	
Spoken pre-test/fm4: Mean duration of repairs	planning	10	.0460	.03307	1.73	18	.005	.053
	control	10	.0220	.02860			.005	.053
Spoken immediate post-test 1/fm4: Mean duration of repairs	planning	10	.0330	.03129	.245	18	.030	.038
	control	10	.0290	.04095			.030	.038
Spoken delayed-test 1/ fm4: Mean duration of repairs	planning	10	.0300	.03432	2.66*	9.153	.006	.051
	control	10	.0010	.00316			.004	.053
Spoken immediate post-test 2/fm4: Mean duration of repairs	planning	10	.0430	.03302	4.11**	9.00	.021	.064
	control	10	.000	.000			.019	.066
Spoken delayed-test 2/ fm4: Mean duration of repairs	planning	10	.0240	.02989	2.11*	10.79	.000	.041
	control	10	.0030	.00949			.000	.042

Significant differences between the groups were identified by delayed test 1 (t (9.15) = 2.66, p= .02), immediate post-test 2 (t (9) = 4.11, p=.003), and delayed test 2 (t (10.79) = 2.11, p=.05). The results suggested that the treatment group were trying to attend to the quality of their production by frequently repairing their utterances. This could have been a natural

outcome of the explicit instructions/treatment they received during the online sessions. This seems in line with the fluency findings discussed in **sections 1.5.9 and 1.5.10** which showed that the treatment group were a bit hesitant in terms of their language use and paused more often than the control group. More consistent results could have been detected if more participants had been recruited in the current study, and perhaps if the learners had a longer treatment duration with their partners.

Summary of the results for oral fluency

To conclude, the treatment group, as compared with the control group, produced significant gains in most of the fluency measures: number of silent pauses (delayed test 1 and delayed test 2), pruned speech rate (delayed test 2), and mean duration of repairs (delayed test 1, post-test 2 and delayed test 2). The findings, therefore, supported the majority of previous planning studies and showed that overall the planning time (e.g., Seyyedi et al., 2013; Thompson, 2014; Yuan & Ellis, 2003) as well as the feedback given to the treatment group had positive consequences for fluency over time. Nevertheless, more consistent findings in terms of fluency were reported in Thompson's (2014) study throughout the immediate and delayed tests. In contrast to my study, Thompson employed tasks which had similar structure (i.e., a family going to buy a pet, a car, shoes in a shop) which perhaps resulted in greater fluency gains over time. In other words, the fact that the learners already knew the structure of the tasks might have freed them to attend to other aspects of their production. The tasks used during the tests in the current study, though all were of the same type (i.e., narrative), each had a different content/structure and required a different variety of lexical items. Another important factor to be considered here when comparing fluency findings with previous studies is the measures used to capture fluency; unlike the current study, which featured multiple and varied measures, most of the previous studies limited their analysis to one or two fluency measures.

All in all, compared with the control group, the treatment group showed an increase over time for almost all the CAF measures in the oral tests. These findings were not surprising; it was expected that the treatment delivered to the planning group during the online sessions would have an impact on learning in some way. Now, though such results can be taken to support Robinson's (2001) cognition hypothesis, it is apparent that the learners were not attending to all the CAF subdimensions at the same level (i.e., significant improvements were not always achieved simultaneously for each subcomponent per test). Thus, in line with previous research and despite the treatment they received during the sessions, the experimental group did focus their attention on some CAF subconstructs to the detriment of others throughout the subsequent spoken tests due to trade-off effects (e.g., Kawauchi, 2005; Mochizuki & Ortega, 2008; Yuan & Ellis, 2003).

4.2 RQ2: Does the oral and written proficiency of intermediate level Syrian learners improve as a result of their longitudinal online interaction with more proficient L2 users?

Unlike the first research question, which compared the subsequent L2 performance of the SLs *across* the planning and the control group, this section will discuss the quantitative data related to research question two which examined learners' L2 production *within* each of the two groups to detect any individual variation. I will therefore begin by presenting and then discussing the descriptive statistics for each of the CAF measures to identify any differences in the performance *within* each group and evaluate their progress across all writing and speaking tests. To this end, paired samples t-tests were carried out to compare the short-term/long-term gains *within* each group between pre- and immediate post-test 1, pre- and delayed test 1, pre- and immediate post-test 2, and finally pre- and delayed test 2.

4.2.1 Complexity measure 1: ratio of clauses per AS-unit

Paired samples t-tests revealed that overall, there was a perceptible increase in the mean scores of the treatment group during the written tests (see **table 32**).

Table 32: ratio of clauses produced by the treatment group during the written tests

Paired Samples Test

Paired Differences

		Mean	SD	95% Confidence Interval of the Difference		t	df
				Lower	Upper		
Pair 1	written pre-test/ratio of clauses per AS-units - written immediate post-test 1/ratio of clauses to AS-units	-.4910	.77174	-1.0430	.06107	-2.012	9
Pair 2	written pre-test/ratio of clauses per AS-units - written delayed-test 1/ ratio of clauses per AS-unit	-.3930	.68865	-.88563	.09963	-1.805	9
Pair 3	written pre-test/ratio of clauses per AS-units - written immediate post-test 2/ratio of clauses per AS-units	-.6610	.78087	-1.2196	-.10240	-2.677*	9
Pair 4	written pre-test/ratio of clauses per AS-units - written delayed-test 2/ ratio of clauses per AS-unit	-.3740	.79210	-.94064	.19264	-1.493	9

That is to say, the learners produced a greater number of clauses per AS-unit during the subsequent written tests as compared with the pre-test, especially in the immediate post-tests (post-test 1 and post-test 2). This indicates that the guided planning and the feedback delivered to the treatment group during the online sessions produced greater short-term gains (compared with long-term gains) with regard to the number of clauses they generated per AS-unit. However, the only significant difference was found when the results of the pre-test and those of immediate post-test 2 were compared; (t (9) =2.67, $p=$.02 sig.(2-tailed)). Longer duration of

treatment could have resulted in more significant differences for the learners within the treatment group across the other tests.

As expected, no progress was detected in the results of the control group regarding this measure throughout the written tests. Rather, compared with the pre-test, there was firm evidence to suggest a significant decrease in the number of clauses the control group produced per AS-unit during delayed test 1 and delayed test 2: (t (9)= 4.80, p= 001 sig.(2-tailed)) and (t (9)= 3.53, p= 006 sig.(2-tailed)) respectively (see **graph 14**). This indicated that there was a trend for the control group to produce simpler language and AS-units than they did in the pre-test as the study proceeded.

Graph 14: complexity measure 1 for the control group in the written tests

As for the speaking tests, more consistent findings were revealed for the treatment group. There was a steady increase in the learners' mean scores throughout the spoken tests (see **table 33**). These results were found to be statistically significant when comparing the mean scores in the pre-test and delayed test 1 (t (9)= 3.07, p= 01 sig.(2-tailed)), the pre-test and immediate post-test 2 (t (9)= 3.35, p= 000 sig.(2-tailed)), and the pre-test and delayed test 2 (t (9)= 4.71, p= 001 sig.(2-tailed)). This indicated an increase in the number of clauses they generated per AS-unit, and hence, one could assume that more syntactically complex sentences were generated

due to the treatment the learners received during the online sessions. The mode of task delivery (i.e., speaking rather than writing) as well as the short practice duration could have led to the non-significant results for immediate post-test 1. In addition, the fact that more significant findings (i.e., greater short-term/long-term effects) were associated with the speaking tests as compared with the writing tests (the delayed post-tests, in particular) was perhaps because in the latter, the learners were typing and seeing their texts on the screen. Influenced by the feedback they received online from their partners, the learners might have tried to attend to different aspects of their production and attempt revisions/self-corrections most often while performing the subsequent written tests; all of which are somehow time-consuming given the limited time the learners had for task completion. This in turn might have resulted in shorter and slightly less syntactically complex AS-units during the immediate written tests.

Table 33: ratio of clauses produced by the treatment group during the spoken tests

Paired Samples Test

Paired Differences

		Mean	Std. Deviation	95% Confidence Interval of the Difference Lower	Upper	t	Df
Pair 1	spoken pre-test/ratio of clause per AS-unit - spoken immediate post-test 1/ratio of clauses per AS-unit	-.26800	.62923	-.71812	.18212	-1.347	9
Pair 2	spoken pre-test/ratio of clause per AS-unit - spoken delayed test 1/ complexity measure1: ratio of clauses per AS-unit	-.82800	.85217	-1.4376	-.21839	-3.073**	9
Pair 3	spoken pre-test/ratio of clause per AS-unit - spoken immediate post-test 2/complexity measure1: ratio of clauses per AS-unit	-.70100	.41410	-.99723	-.40477	-5.353***	9
Pair 4	spoken pre-test/ratio of clause per AS-unit - spoken delayed test 2/ ratio of clause per AS-unit	-.85700	.57424	-1.2677	-.44621	-4.719***	9

The control group, on the other hand, did not show any progress in terms of the number of clauses produced per AS-unit throughout the immediate and delayed oral post-tests compared with the pre-test. There was a significant decrease, however, when the findings related to this measure during the pre-test and immediate post-test 1 were compared; (t (9)= 2.60, p= 02 sig.(2-tailed)). This could possibly be traced back to the direct effects of the text-based chatting sessions where the learners might have noticed a gap in their L2 production as they were interacting with more proficient L2 users and hence preferred to produce shorter and simpler utterances.

4.2.2 Complexity measure 2: average length of AS-unit

Paired samples t-tests uncovered very similar results to those found for the previous measure, in terms of the number of words the learners produced per AS-unit. The treatment group produced relatively longer AS-units throughout the post- and delayed tests as compared with the pre-test during the writing tests. However, the results only turned out to be statistically significant for immediate post-test 2; (t (9)= 3.67, p= 005 sig.(2-tailed)) (see **graph 15**). This could have been due to the instant effects of the voice chat sessions where the learners had some time to verbally practise their L2 with their partners. Yet it may have been that these effects as well as the treatment duration were not sufficient to result in more significant progress as manifested in the other tests. Hence, short-term gains regarding this measure during the writing tests were only noticeable immediately following the voice chat sessions, and did not translate into long-term gains. As previously mentioned, one explanation for the lack of significant differences during the other tests could be because of the mode of task delivery; when learners are writing/typing, their attention might be diverted as they need to focus on multiple aspects of their production. This could have negatively affected the length of their AS-units.

Graph 15: complexity measure 2 for the planning group in the written tests
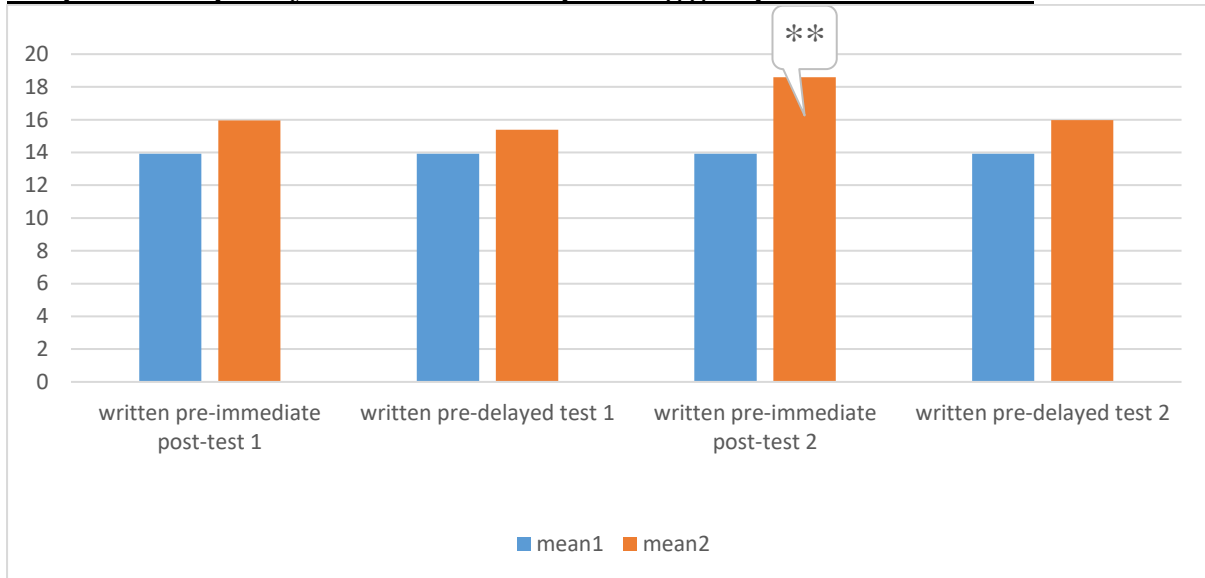
Regarding the control group, their mean scores indicated that they tended to produce shorter AS-units in the post- and delayed tests as compared with the pre-test (see **table 34**). Differences within the control group related to this measure were all found to be statistically significant, i.e., there was a significant decrease in the number of words they produced per AS-unit during the written tests: immediate post-test 1(t (9)= 3.78, p= .004 sig.(2-tailed)), delayed test 1 (t (9)= 4.41, p= .002 sig.(2-tailed)), immediate post-test 2 (t (9)= 2.54, p= .03 sig.(2-tailed)), and delayed test 2 (t (9)= 2.68, p= .02 sig.(2-tailed)). This was expected given the fact that they had not received any kind of treatment or instruction during the online sessions. As the study proceeded and as learners were engaged in more cognitively challenging activities compared with the pre-test, the learners might have felt less confident to produce long sentences and therefore preferred to mainly focus on fulfilling the task's goal.

Table 34: words per AS unit produced by the control group during the written tests

Paired Samples Test

Paired Differences

		Mean	Std. Deviation	95% Confidence Interval of the Difference		t	df
				Lower	Upper		
Pair 1	written pre-test/complexity measure 2 - written immediate post-test 1/ complexity measure 2	4.45700	3.72682	1.79099	7.12301	3.782 **	9
Pair 2	written pre-test/complexity measure 2 - written delayed test 1/complexity measure 2	4.57200	3.27377	2.23009	6.91391	4.416 **	9
Pair 3	written pre-test/complexity measure 2 - written immediate post-test 2/ complexity measure 2	2.88300	3.57752	.32379	5.44221	2.548 *	9
Pair 4	written pre-test/complexity measure 2 - written delayed test 2/complexity measure 2	3.69900	4.35524	.58245	6.81355	2.685 *	9

As for the speaking tests, the treatment group also produced more words per AS-unit as compared with the pre-test. Consistent with the findings of the previous measure, this difference within the group was found to be statistically significant during delayed test 1 (*t* (9)= 2.81, *p*= .02 sig.(2-tailed)), immediate post-test 2 (*t* (9)= 4.34, *p*= .002 sig.(2-tailed)), and delayed test 2 (*t* (9)= 3.06, *p*= .01 sig.(2-tailed)) (see **graph 16**). Thus, short-term and long-term gains were detected for this measure during the speaking tests. The lack of significant differences between the pre-test and immediate post-test 1 could on the one hand be due to the effects of text-based chatting sessions and the fact that the learners were speaking at a slower pace to attend to their errors. It might also be traced back to the short treatment duration; the learners perhaps needed more time to practise and achieve significantly better results.

Graph 16: complexity measure 2 for the planning group in the oral tests
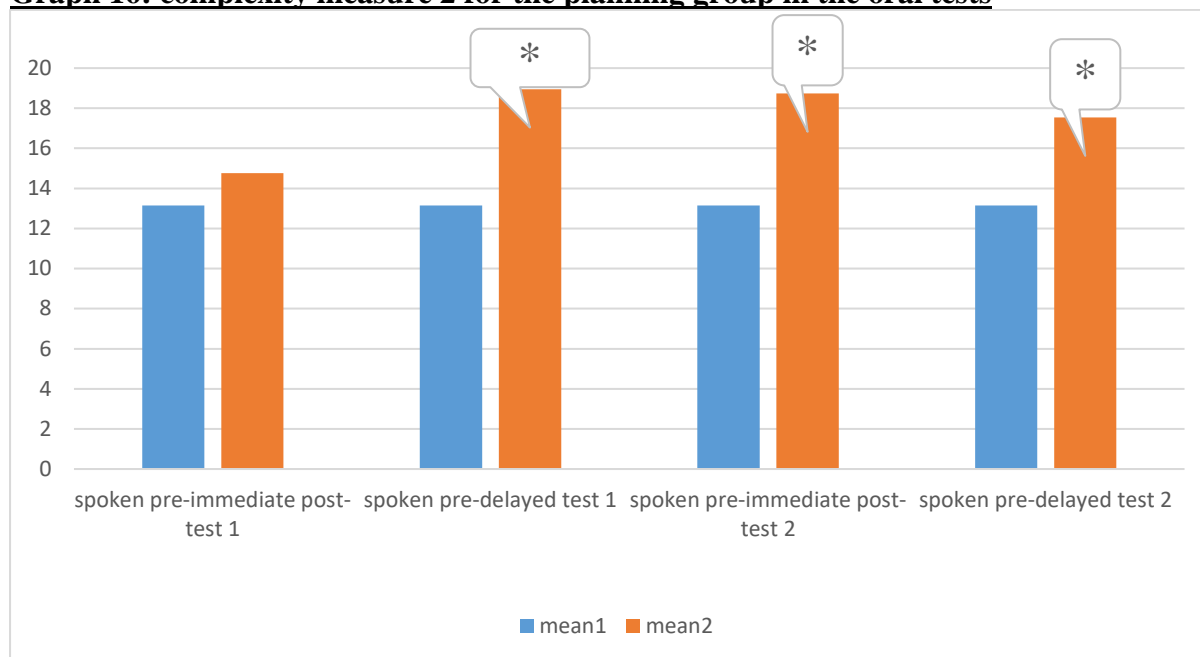
The control group unsurprisingly showed no progress regarding this measure; they did not demonstrate any improvement in terms of the number of words they produced per AS-unit. On the contrary, when comparing the mean scores between the pre-test (M=12.32, SD= 1.98) and immediate post-test1 (M=10.65, SD=1.26), they appeared in line with the findings of the previous measure and indicated a decrease in the average length of AS-units between the two tests as well. This difference was found to be statistically significant (t (9)= 2.29, p= .04 sig.(2-tailed)). Likewise, the fact that the control group did not receive any kind of treatment or instruction during the online sessions was clearly reflected in the number of words they produced later throughout the immediate and delayed tests, which were more challenging compared with the pre-test. That is, the learners might have preferred to produce short sentences and mainly focus on fulfilling the task's goal.

4.2.3 Complexity measure 3: number of complex grammatical structures

Paired samples t-tests revealed that the treatment group produced a greater number of complex grammatical structures (relative clauses, in particular) during the written post- and delayed

tests. The only statistically significant difference was found, however, in immediate post-test 2 where the number of relative clauses (indicated by mean scores; see **table 35**) the learners generated, as compared with the pre-test, was nearly tripled (t (9)= 3.03, p= .01 sig.(2-tailed)). This could possibly be traced back to the instant effects of the voice chat sessions; the immediacy of turn-taking during the voice chat mode could have enabled the more proficient L2 users to instantly comment on the learners' language use. This, however, was not the case during the text-based chat because of the overlap between the turns and participants' involvement with writing and revising their own messages. Additionally, as the voice chat sessions commenced, the learners had already built a rapport with their partners and hence the latter might have felt that it would be more convenient and perhaps less intimidating for the learners to explicitly address their errors and trigger them to frequently incorporate these complex structures into their L2 production. It might also be rational to claim that the different contents of the tasks could have influenced learners' production of complex grammatical structures, i.e., different contents/topics might require a varied use of these forms. It might be worth noting that the lack of short-term/long-term gains across the other tests could be traced back to the complexity of these forms as well as the frequency and the duration of the online sessions.

Table 35: complex grammatical structures the treatment group produced during the written tests

Paired Samples Test

Paired Differences

| | | Mean | Std. Deviation | 95% Confidence Interval of the Difference | | t | df |
				Lower	Upper		
Pair 1	written pre-test/complexity measure 3 - written immediate post-test 1/ complexity measure 3	-.400	.843	-1.003	.203	-1.500	9
Pair 2	written pre-test/complexity measure 3 - written delayed test 1/complexity measure 3	-.500	1.179	-1.343	.343	-1.342	9
Pair 3	written pre-test/complexity measure 3 - written immediate post-test 2/ complexity measure 3	-1.800	1.874	-3.140	-.460	-3.038**	9
Pair 4	written pre-test/complexity measure 3 - written delayed test 2/complexity measure 3	-.500	1.269	-1.408	.408	-1.246	9

With regard to the speaking tests, the results were less consistent than the writing tests; there was no progress in terms of the number of relative clauses generated by the treatment group as compared with the pre-test (M=80. SD=.789). To put it differently, the learners tended to produce fewer relative clauses as the study proceeded. There was a slight increase though during delayed test 2 (M=90, SD=1.19), but no significant differences were found. It could be plausible to claim that the complexity of the forms, the mode of task delivery, and the fact that the learners only had one minute to achieve the task goal might have prevented them from integrating these forms/complex structures into their discourse very often during the tests, especially if the learners were attending to different aspects of their production at the same time. That is to say, if the learners had not been pressured by addressing the task's goal within a specific time limit, they might have made greater gains and integrated more relative clauses into their L2 production. Unsurprisingly, there was a decrease in the number of relative clauses

the control group produced, yet no significant change was captured within this group during the written and spoken tests.

4.2.4 Complexity measure 4: syntactic variety

Paired samples t-tests did not reveal any significant improvement regarding this measure within the two groups during either the writing or the speaking tests. Marginal gains were achieved, however, by the treatment group for the number of different verb forms (in terms of tense, voice, and modality) they produced compared with the pre-tests in written delayed test 2: written pre-test (M=4.10, SD=1.37) and written delayed test 2 (M=4.60, SD=.699); and spoken delayed test 2: spoken pre-test (M=3.60, SD=.966) and spoken delayed test 2 (M=4.20, SD=.919). The lack of significant differences related to this measure, particularly within the treatment group, could be due to the nature and type of the tasks used during the tests: narrative tasks were used and hence the learners might have felt that using either simple past or simple present tenses would be most appropriate to achieve the task's goal. Additionally, based on my previous teaching experience and familiarity with what EFL learners (especially in Syria) often consider as complex L2 structures, SLs might have deliberately chosen not to integrate the passive voice into their production during the tests. As stated earlier, they could have preferred not to take risks or use forms that are cognitively demanding, and which normally need more time to be processed and/or incorporated into their L2 production. Perhaps including explicit instructions and guidance to integrate such verb forms could have resulted in a greater variety of verb forms throughout the tests.

4.2.5 Complexity measure 5: lexical variety

The findings revealed that less lexically varied texts were produced by the treatment group during the written post- and delayed tests as compared with the pre-test. A very slight increase

was detected in immediate post-test 1. Possibly this result was influenced by the text-based chat sessions and the fact that the learners were seeing their partners' messages on the screen. Perhaps they were noticing new lexical items in their partners' production and tried to employ these items during the test.

Table 36: lexical variety of the treatment group's production per written test

Paired Samples Test

Paired Differences

		Mean	Std. Deviation	95% Confidence Interval of the Difference		t	df
				Lower	Upper		
Pair 1	written pre-test/complexity measure5 - written immediate post-test 1/complexity measure5	-2.333	27.806	-22.224	17.558	-.265	9
Pair 2	written pre-test/complexity measure5 - written delayed test 1/complexity measure5	5.733	18.401	-7.4308	18.896	.985	9
Pair 3	written pre-test/complexity measure5 - written immediate post-test 2/complexity measure5	11.677	14.965	.97153	22.382	2.467*	9
Pair 4	written pre-test/complexity measure5 - written delayed test 2/complexity measure5	9.263	21.161	-5.8748	24.400	1.384	9

A significant decrease, however, was found when the mean scores (see **table 36**) during the pre-test and immediate post-test 2 were compared; ($t(9)=2.46$, $p=.03$ (sig.2-tailed)). This could possibly be due to the intervention of the voice chat sessions, where learners were just listening (without seeing the conversation on the screen), and probably not understanding every single

word their partners were saying. Therefore, it seems highly unlikely that learners would be able to integrate these words into their own utterances. Another possible explanation might be that the different task type/content the learners were asked to work on had not given them the chance to reuse these words throughout the tests. Additionally, the lack of long-term gains could be because the learners did not have the opportunity to practise their L2 and reproduce the newly learned vocabulary in their contexts.

As for the speaking tests, similar results were obtained for the treatment group: learners' production during the post- and delayed tests was slightly less lexically varied than for the pre-test. The only exception to this was for immediate post-test 2 where the learners seemed to achieve marginal gains: pre-test (M=4.41, SD=14.39) and immediate post-test 2 (M=43.75, SD=12.48). The opportunity the learners had to practise their oral skills with more proficient L2 users during the voice chat sessions could have led to this increase in the variety of the lexical items they generated during the following test. Yet neither significant differences nor short-term/long-term gains were detected within the treatment group for this measure. Again, this lack of significant gains could be traced back to the different content/contexts of the tasks as well as the time pressure the learners were operating under, especially when performing the speaking tests.

Unsurprisingly, no progress was detected for the control group in terms of the lexical variety of their written and spoken tests. In fact, a significant decrease was found when comparing the learners' results during the written pre-test and delayed test 1 (t (9) =2.21, p=.05 sig.(2-tailed)).

4.2.6 Complexity measure 6: lexical appropriacy

Paired samples-tests showed that compared with the written pre-test (M=1.00, SD=1.05), slightly higher mean scores (i.e., more inappropriate lexical items in terms of meaning) were generated by the learners working under the treatment group immediately following the

chatting sessions (written immediate post-test 1 and written immediate post-test 2), (M=1.50, SD=1.26) and (M=1.40, SD=.843) respectively. This could be attributed to the fact that the learners were learning new expressions/words from their partners throughout the online sessions and might have felt eager to use them later during the written tests, especially as they had some time to review and edit their narratives. However, as the learners were asked to work on different task contexts, not all of their attempts to reuse new learned lexical items may have been successful.

More inappropriate lexical items were also employed by the treatment group across the spoken post- and delayed tests. During immediate post-test 1, in particular, the learners had the highest number of inaccurate lexical items (M=.70, SD=.949) as compared with the pre-test (M=.10, SD=.316). Again, the affordances of text-based chat interaction and the prominence of the messages on the screen may have made these new lexical items more noticeable for the learners who in turn could have tried to use these items in the following test. In addition, the mode of task delivery could have also affected the results for this measure. To put it simply, during the speaking tests the learners spoke spontaneously and possibly used the first words/expressions that came to their mind without paying much attention to whether or not these items were accurate/appropriate (in terms of meaning) for the given context. No statistically significant difference was found for the treatment group in either the written or spoken tests regarding this measure. No change or significant differences were detected for the control group during the tests either.

4.2.7 <u>Accuracy measure 1: percentage of error-free clauses</u>

The treatment group improved in terms of the percentage of error-free clauses they generated during the post- and delayed tests (except for immediate post-test 2). Although the learners had perceptibly improved in immediate post-test 1, the improvements were not statistically

significant. One explanation for the non-significant findings could be that by the time they did the first immediate test, learners had been practising their L2 for only a short period of time; and therefore might not have had the chance to attend to all the problematic issues they had in their writing.

The learners maintained their progress over delayed test 1; i.e., as the learners had more practice at becoming acquainted with the nature of the test as well as the type of task, they tended to produce more error-free clauses compared with the previous tests. This difference was found to be statistically significant (t (9)= 3.45, p= .007 (sig.2-tailed)) (see **graph 17**). Hence, there is firm evidence to suggest that the learners achieve long-term gains with regard to this measure and benefitted from the feedback they received from their partners during the text-based chatting sessions.

Graph 17: Accuracy measure 1 for the planning group

Another significant difference was detected in immediate post-test 2; however, this time the learners generated significantly fewer error-free clauses than in the pre-test (t (9)= 2.45, p= .03 (sig.2-tailed)). A rational explanation could be that the learners were influenced by the voice chat sessions where their foremost aims were to focus on meaning and maintain the flow of the conversation with their L2 partners, paying less attention to how accurate their production was.

It is worth noting though that the learners improved their scores in delayed test 2 and produced more error-free clauses than the pre-test, although the differences were non-significant. Hence, significant long-term accuracy gains were found only in delayed test 1.

As for the speaking tests, the treatment group produced more error-free clauses in immediate post-test 1 and delayed test 2 compared with the pre-test, and very marginal decreases resulted from delayed test 1 and immediate post-test 2; however, none of the results were statistically significant. The inconsistent progress and the lack of significant findings with regard to this measure throughout the speaking tests could be stemmed from the mode of task/test delivery. As stated earlier, the time pressure might have limited the learners' attentional resources, and consequently they might have not been able to keep track of and attend to every aspect of their L2 production.

As expected, not much progress was noticed for the control group during the writing tests; there was a very minor increase during immediate post-test 1 and delayed test 1 compared with the pre-test. Possibly this was due to the saliency of the L2 production as exchanges appeared on learners' screens and remained there for some time during the text-based chat sessions, which may have improved the potential for awareness raising, noticing, and acquisition. A noticeable difference was found, however, during immediate post-test 2, where the learners produced significantly fewer error-free clauses than the pre-test: (t (9)= 2.45, p= .03 (sig.2-tailed)). No progress or significant results were uncovered within the control group during the speaking tests.

4.2.8 Accuracy measure 2: percentage of errors per test

Paired samples t-tests regarding the percentage of errors per test were consistent with those found for the previous accuracy measure; overall, the treatment group had lower mean scores (i.e., generated a smaller number of errors) during the written post- and delayed tests compared

with the pre-test. An exception concerned immediate post-test 2, where there was an increase in the percentage of errors produced per test (see **table 37**); however, this increase was not found to be significant. This result could have been affected by the voice chat sessions and learners' involvement with focusing on delivering meaning, given the limited time they had to address the task's goal.

Table 37: percentage of errors produced by the planning group during the written tests

Paired Samples Test

Paired Differences

		Mean	Std. Deviation	95% Confidence Interval of the Difference		t	df
				Lower	*Upper*		
Pair 1	written pre-test/accuracy measure 2 - written immediate post-test 1/accuracy measure 2	.8960	5.178	-2.808	4.600	.547	9
Pair 2	written pre-test/accuracy measure 2 - written delayed test 1/accuracy measure 2	4.059	3.607	1.478	6.639	3.558**	9
Pair 3	written pre-test/accuracy measure 2 - written immediate post-test 2/accuracy measure 2	-4.014	5.889	-8.226	.1988	-2.155	9
Pair 4	written pre-test/accuracy measure 2 - written delayed test 2/accuracy measure 2	1.325	4.536	-1.920	4.570	.924	9

There was, however, strong evidence to suggest long-term gains with regard to this measure in delayed test 1 (t (9)= 3.55, p= .006 (sig.2-tailed)). Again, this was in line with the findings of the previous measure discussed earlier, where a greater number of error-free clauses was

detected. Hence, apart from the results of immediate post-test 2, there were short-term and long-term gains for this measure; yet not all were found to be statistically significant.

Compared with the spoken pre-test, there was also a trend for the treatment group to produce fewer errors in immediate post-test 1 and this was maintained for immediate post-test 2 and delayed test 2; a marginal increase in the number of errors was found in delayed test 1. None of the results were statistically significant which was perhaps due to the short-treatment duration. Therefore, a longer, more frequent, and more focused treatment (that involved addressing specific errors in learners' production, such as tense and article use) could have resulted in greater short-term and long-term gains.

On the contrary, as the study proceeded, the control group produced more errors than they did in the written pre-test (M=10.36, SD=5.41). There was a great increase in the number of errors produced by the learners in immediate post-test 2 (M=16.61, SD=7.16); this difference was significant (t (9)= 2.69, p= .02 (sig.2-tailed)). This was expected since learners working under the control group were not receiving any kind of treatment or feedback during the online sessions related to the problematic issues they had in their L2 production. Furthermore, they were pressured by the limited time they had to create their oral narratives. As anticipated, due to the demands of the mode of interaction and task delivery, no gains were made by the control group during the speaking tests.

4.2.9 Fluency measure 1 for the written tests: rate of production

The findings showed that during the written post- and delayed tests, the treatment group produced more words than they did in the pre-test, given the time they spent on the task. A very slight decrease was found in terms of the rate of learners' production in written immediate post-test 1; a plausible justification could be that the learners were affected by the slow pace of the interaction during the text-based chatting sessions. It could also be because the test used

the written mode, and hence the lack of progress and short-term gains for this measure might be traced back to the learners frequently revising and editing their narrative before submitting the final version.

The treatment group started to show progress and make short-term/long-term gains in the other tests, especially following the voice chat sessions, yet only achieving significant results during delayed test 2 (t (9)= 3.10, p= .01 (sig.2-tailed)) (see **graph 18**). Again, the written mode of the task (during the tests), as well as the treatment delivered throughout the online sessions could perhaps have made the learners more attentive to their writing, trying to address multiple aspects of their L2 production.

Graph 18: fluency measure 1 for the planning group in the written tests
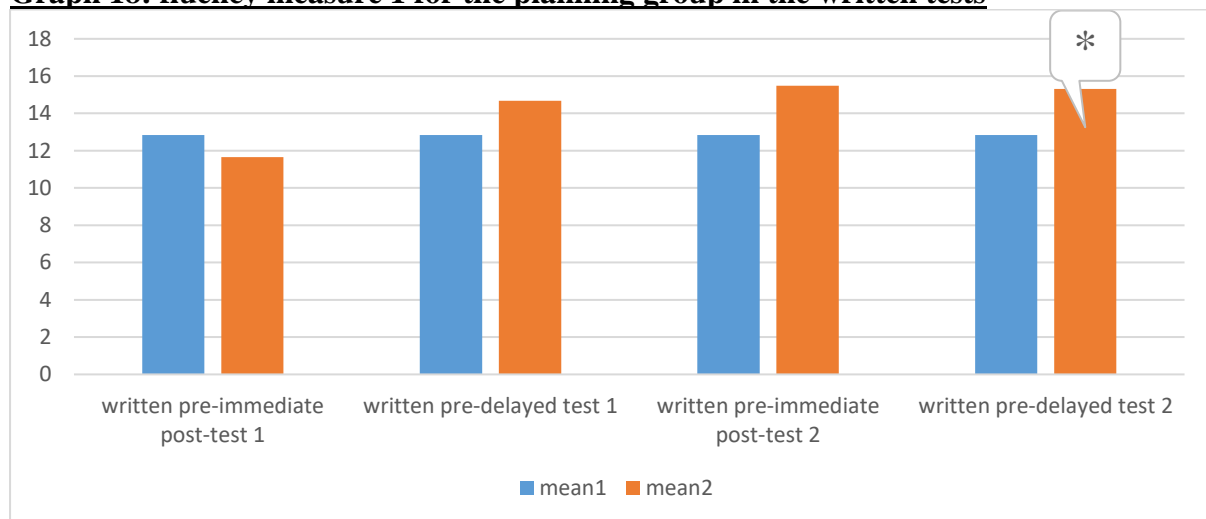
As for the control group, the findings revealed that they managed to achieve gradual progress in terms of the number of words they generated given the time they spent on the task during the writing tests. When comparing the learners' results in the written pre-test (M=9.69, SD=2.72) and those of delayed test 2 (M=12.95, SD=2.90), a statistically significant difference was uncovered (t (9)= 3.99, p= .003 (sig.2-tailed)). Different reasons could be advanced to interpret learners' development and the above-mentioned significant result. The control group was also given the chance to practise their L2 with more proficient L2 users, and therefore it might be logical to expect to see less disfluency in their production. This does not necessarily

mean, however, that this production was complex and accurate as well. To put it differently, since the control group was not receiving any feedback or treatment during the online sessions, they might have been primarily focusing on delivering meaning and achieving the task goal without spending much time on editing their production. No significant results were captured during the other writing tests. It is worth noting that this was the only measure used to gauge the fluency of learners' production during the written tests. As illustrated in the following sections, a different set of fluency measures were employed for the spoken tests.

4.2.10 <u>Fluency measure 1 for the spoken tests: number of silent pauses</u>

Paired samples t-tests showed that the treatment group made fewer silent pauses while orally narrating the story during the speaking tests compared with the pre-test (M=.36, SD=.08). One exception was found for immediate post-test 1 (M=.39, SD=.04), where the learners produced more silent pauses than in all the other tests. This could be considered as a plausible outcome of the text based online sessions where the learners noticed the gaps they had in their L2 as they were viewing their output and that of their partners on the screen and receiving feedback on their problematic L2 production. This in turn could have made the learners more careful about the language they produced, particularly because they were orally narrating the story and thus did not have much time to think of what they wanted to say. Therefore, they might have chosen to follow a strategy of taking short breaks to think of their sentence structure, word choice, etc., and lessen the opportunity to make more errors. There were, however, no significant differences within the treatment group regarding this measure. As mentioned earlier, due to the lack of access to keystroke logging software, it was not possible to use these measures to assess learners' fluency during the writing tests.

The control group also showed progress in terms of the number of silent pauses produced per minute; they made fewer pauses compared with the speaking pre-test (M=.44, SD=.06). A

significant decrease was uncovered for immediate post-test 2 (t (9)= 3.16, p= .01 (sig.2-tailed)) and delayed test 2 (t (9)= 3.26, p= .01 (sig.2-tailed)) (see **graph 19**). Again, the results were probably influenced by the voice chat sessions; after all, learners were practising their L2 with more proficient L2 users which could have resulted in more positive outcomes in terms of the number of silent pauses generated per minute.

Graph 19: fluency measure 1 for the control group in the spoken tests
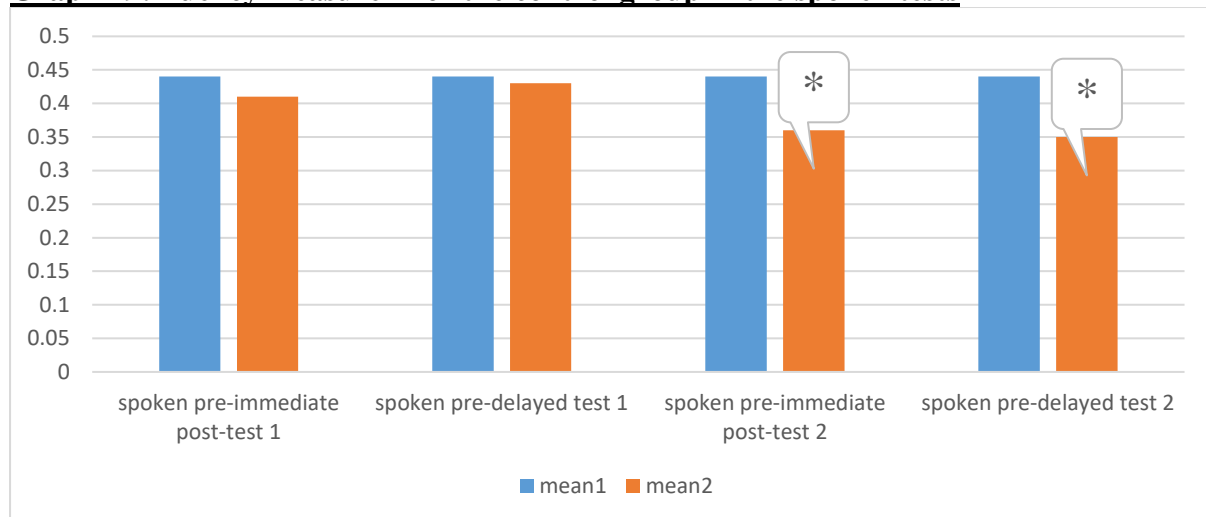
4.2.11 Fluency measure 2 for the spoken tests: number of filled pauses

The treatment group, possibly influenced by the speaking practice they experienced during the voice chat sessions, produced a lesser number of filled pauses during spoken immediate post-test 2 (M= .167, SD= .142) and delayed test 2 (M=.124, SD=.105) as compared with the pre-test (M=.169, SD= .138). No positive change was noticed during the earlier tests and none of the results for this measure throughout the tests were statistically significant. Perhaps this was caused by the learners' attempts to focus on different aspects of their production, or maybe because the practice they had was insufficient for them to show equal progress regarding all the measures. No change or significant differences were found for the control group, either.

4.2.12 <u>Fluency measure 3 for the spoken tests: pruned speech rate</u>

As the study proceeded, the treatment group produced more syllables of pruned speech (i.e., number of meaningful syllables produced excluding repetitions, false starts, and so on) per minute than they did in the pre-test (M=125.19, SD=20). Not much change was noticed after the text-based chat sessions, possibly due to the slow pace of interaction during this mode. Given the increased number of pauses generated by the treatment group (as discussed earlier), the learners might have been affected by the written chat mode. In other words, they were trying to pay more attention to their L2 output, taking more time to carefully think of the structure of their utterances, which in turn could have negatively influenced the number of syllables they produced per test. The finding that there was no change in the number of syllables the learners produced in immediate post-test 1 and delayed test 1 could also be traced back to the lack of L2 speaking practice; at first, the learners might have preferred to avoid taking the risk of making errors, and therefore tended to produce a concise oral narration of the pictures. However, following the voice chat sessions, the learners started to generate a greater number of meaningful syllables, as seen by the results of immediate post-test 2 (M=136.5, SD=21.38). The learners also maintained this progress in delayed test 2 (M=155, SD=25.99), and eventually managed to achieve statistically significant long-term gains (t (9)= 4.02, p=.003 (sig.2-tailed)) (see **graph 20**). One could presume that the voice chat sessions, which involved giving the learners the opportunity to speak and practise their L2 with more proficient L2 users, at least partially account for these results, giving the learners the confidence to produce more language. Perhaps more speaking practice and a longer treatment duration could have resulted in greater short-term and long-term gains. The control group, on the other hand, also showed some progress regarding this measure after the voice chat intervention, yet no significant gains resulted.

Graph 20: fluency measure 3 for the planning group in the spoken tests
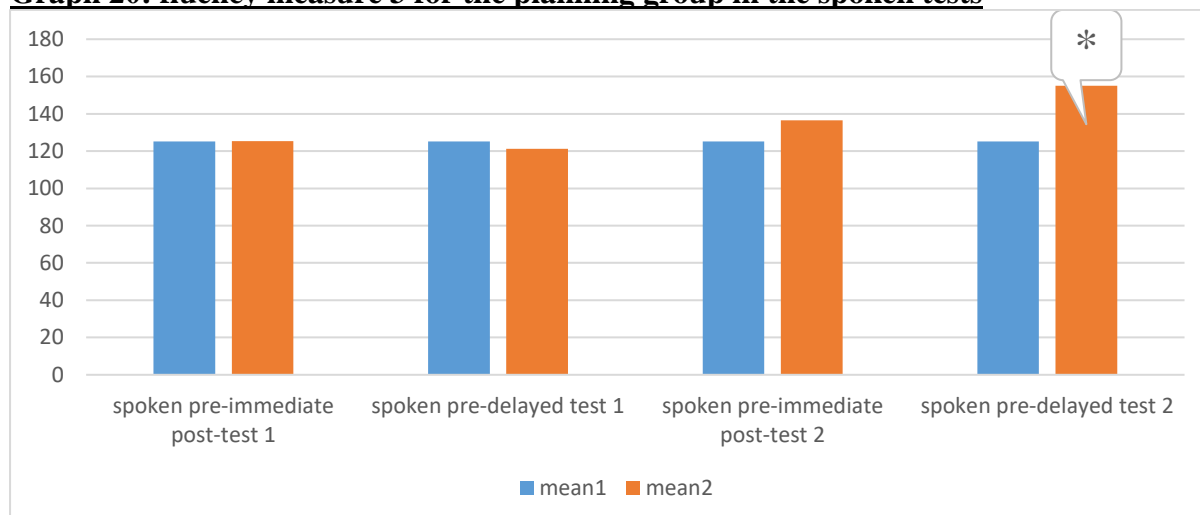
4.2.13 Fluency measure 4 for the spoken tests: mean duration of repairs

Paired samples t-tests showed that the treatment group frequently attempted self-repair during the tests. The highest mean score was found during the pre-test (M=.046, SD=.033), where the learners tended to repair their production more often than they did during the following tests. This could possibly be due to the pressure the learners were experiencing at the beginning of the study; although the learners were not aware that they were doing a test and they were frequently told that their narrative would not be assessed as right or wrong, controlling their stress levels was beyond the control of the researcher. Another very similar result was revealed for immediate post-test 2 (M=.043, SD=.033) which could be traced back to the direct effects of the voice chat sessions; although the learners produced more syllables per test (as shown earlier), they were perhaps very concerned about their L2 output (because of the guided planning and the feedback they received in the sessions), and accordingly more repairs were found in this test compared with the other tests. A slight decrease, however, resulted after the text-based chat sessions, as seen in immediate post-test 1 (M=.033, SD=.031) and delayed test 1 (M=.030, SD=.034). Influenced by these sessions and the saliency of the messages on their screens, the learners might have started to feel more confident about their production. A further

decrease was found in delayed test 2 (M=.024, SD=.029), which turned out to be statistically significant when compared with the pre-test: (t (9)= 2.65, p=.02 (sig.2-tailed)) (see **graph 21**). It could be presumed that by the time the learners did this test, they had already had some practice in terms of their speaking skills and built up their L2 knowledge and/or confidence; therefore, there was no need to attempt self-repair as much as throughout the previous tests.

Graph 21: fluency measure 4 for the planning group in the spoken tests
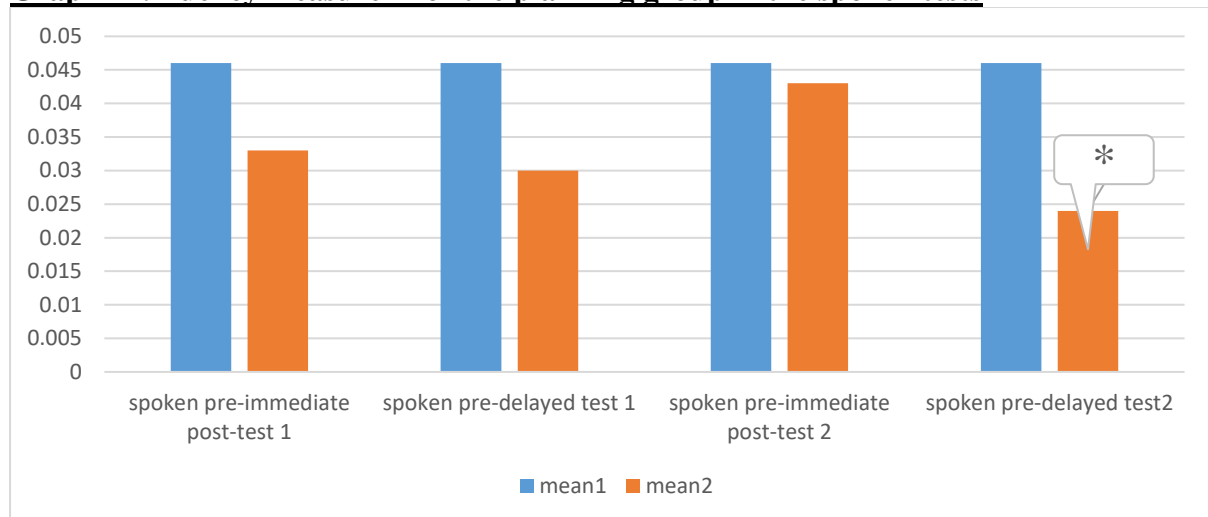
As for the control group, few repair attempts were made by the learners and all the results, except those during immediate post-test 1, were found to be statically significant when compared with the pre-test: pre-test/delayed test1 (t (9)= 2.40, p=.04 (sig.2-tailed)), pre-test/immediate posttest 2 (t (9)= 2.43, p=.03 (sig.2-tailed)), and pre-test/delayed test 2 (t (9)= 2.23, p=.05 (sig.2-tailed)). These significant results indicated that as the learners were not receiving any kind of treatment during the online sessions, they perhaps were not very concerned about the quality of their production and instead were only focusing on addressing the task's goal.

Summary of the Quantitative Findings:

All in all, the pre-planning condition as well as the explicit instructions/feedback given to the treatment group during the online sessions seemed to have positive effects on the learners' L2 development. Due to the variety of CAF measures (both general and specific) employed in the current study to gauge L2 progress, different findings were revealed. For example, fewer short-term/long-term gains were uncovered for lexical variety. Perhaps this could be due to the different contexts/types of tasks the learners worked on during the online sessions, which in turn might have inhibited learners' implementation of newly learned lexical items. Besides, it is worth mentioning that coming across a new word/expression does not necessarily mean fully understanding it and having the ability to use it correctly in future contexts (Nation, 2015). A carefully designed and focused task that frequently triggered the use of a particular set of lexical items in different contexts could have resulted in a significant acquisition of these items.

The findings also revealed that the text-based chat mode exerted a greater positive influence on the accuracy of learners' narratives in the written and spoken tests. Hence, the feedback delivered to the learners during the text-based chat sessions, the saliency of L2 production, and the slow pace of the interaction, which gave the learners the opportunity to review and edit their messages before sending them, were all found to be more effective than the audio chat to improve learners' accuracy. To put it another way, during text-chat conditions, learners can pay more attention to the accuracy and complexity of the language they produce, yet in oral tasks they might be so preoccupied with aspects such as pronunciation and time pressure that there is little room for improvement in accuracy and complexity. This finding has important pedagogical implications for teachers who might take advantage of the affordances of text-based chat mode to scaffold the learners to attend to specific gaps in their production.

Finally, when we examine the fluency gains of learners' production throughout the tests, results were mixed. This could be traced back to the lack of L2 speaking practice in their foreign

language context and the short duration of the online sessions. More significant gains, however, were detected following the voice chat sessions, which indicated a positive influence of these oral sessions on learners' fluency.

Having discussed the quantitative results for the first two questions, the following chapter will describe the qualitative results for the third, fourth, and fifth research questions.

5. Results & Discussion (2)

The purpose of this chapter is to report the qualitative results of this study. The chapter will be divided into three parts. The first part will qualitatively examine the strategies followed by the SLs to prepare for their task performance prior to/during the online sessions and scrutinise the rationale behind their planning choices. The factors that impacted learners' L2 development across different modes of online interaction as well as the participants' perceptions towards the online experience will be presented in the second and third parts respectively.

Part 2: Qualitative Findings

During the online sessions, a particular task sequence was followed with the aim of gradually scaffolding the learners to work with their partners on simple to more complex tasks. The first two tasks of the text-based chat (spot the difference and information-gap) were meant to act as an ice breaker, so that the cognitive load the participants might experience, as they did not get to know each other prior to the beginning of the study, would be minimised. Then more complex and cognitively demanding tasks (narrative and decision-making) were employed in the following sessions. As the voice chat sessions started, the learners had already built a rapport with their partners. Nevertheless, it was assumed that the complexity of the verbal mode, being more instantaneous and more akin to face-to-face conversation than the text-based chat, would increase the cognitive load on the part of the learners. Therefore, and in an attempt to ease the cognitive burden on the learners, they were asked to work on another spot-the-difference task in the first session of voice chat, followed by a narrative task and four other decision-making tasks; all were sequenced based on their content and cognitive demands (i.e., simple to complex). Planning notes, transcripts of the online conversations of four SLs (three from the planning group and one from the control group), the recall interviews conducted with

them afterwards as well as their weekly reports were therefore utilized as the main data sources to introduce the qualitative findings.

Due to time constraints as well as space constraints in the thesis, it was not possible to fully analyse the data sources of all the SLs who took part in the current project. Instead, a case study approach was adopted; only four cases were qualitatively analysed in detail (three learners from the treatment group and one from the control group). As mentioned in (section **3.3.8.2**), the choice of the case studies was based on the progress achieved by the learners in terms of their language use (i.e., whether they produced more complex, accurate, and fluent language as the sessions continued), the nature of the feedback delivered by their partners (i.e., implicit and/or explicit), the rapport (or lack of rapport) between the learners and their interlocutors, and whether or not the learners exhibited a different behaviour throughout the sessions, particularly in terms of their motivation/anxiety level. Before moving on to present the results for the third research question, a quick reminder of each case study will be provided below.

- Majd, who was preparing to go to Poland for an IT course, was assigned to work with a NS under the guided planning condition. The findings disclosed that Majd was producing more complex, accurate, and fluent output as the sessions proceeded.

- Lara was also assigned to work in the treatment group and received guided instructions. Qualitative analysis of the chat logs showed that Lara developed a very good rapport with her partner, Kate, and was gradually progressing throughout the sessions, becoming less anxious and hence more confident about her language use.

- Ameen was also working in the treatment group and preparing to go to Poland to attend an IT course in English. Unlike Majd and Lara, Ameen was not showing much progress in terms of his language use. Having analysed his chat logs as well as responses during the recall interviews, lack of rapport between Ameen and his partner, Rami, was very

evident throughout the online sessions. Furthermore, Ameen also seemed a bit reticent to express his thoughts during the interview although he was using his L1.

- Rima, on the other hand, was assigned to work in the control group and did only receive instructions related to task's goal/content prior to the online performance. Qualitative data from her chat logs disclosed that Rima was getting more motivated about improving her English as the sessions were proceeding and was asking questions about her language use. She was also showing some progress and was writing/speaking more compared with the previous sessions.

5.1 RQ3: What strategies did the Syrian learners use when planning for different task types across different modes of online interaction (text VS voice chat)?

I began the analysis by examining the notes taken by the learners during the planning time they had been offered prior to each chatting session, and the language they produced as they were interacting online with their partners. Learners' metacognitive responses in the interviews with regard to the cognitive processes they were engaged in before and during the online interaction were also analysed. Answering this interview question would provide fruitful insights into the strategies learners used while planning for different task types across different modes of online interaction, whether these strategies changed in subsequent performances and why. Such knowledge would provide useful information as to the way learners attend to different aspects of their L2 production over time. Hence, the planning choices made by the learners (who were chosen as case studies) per mode of interaction as well as their comments on these choices will be presented and fully discussed in the following.

5.1.1 <u>Planning strategies for different task type before and during the text chat sessions</u>

The findings revealed that whilst planning for the online sessions, the learners differed in terms of the aspect(s) of their language they chose to focus on. The recall interviews also revealed that task type, mode of interaction, L1 status of the more proficient L2 speaker, and individual learner differences substantially influenced how learners perceived the benefits of planning time and the strategies they used prior to and during the actual online performance. In line with Ortega's (2005) argument, some learners were primarily concerned with producing accurate output; and therefore, they appreciated the time they were given to plan their online performance and pay more attention to form. However, other learners had a natural inclination to convey meaning when writing/speaking in the L2, and therefore, they preferred to start the session with their partners immediately. Hence, the planning strategies adopted prior to and during the online sessions (as reported by the learners, and which also include my analysis of the other data sources, such as the video recordings) will be presented per case study in the following paragraphs

5.1.1.1 Case Study 1 (Majd):

Starting with Majd, as illustrated in **table 37** below, he was trying to attend to different aspects of his language while planning different types of tasks for the text-based chat sessions. Having worked on the spot-the-difference task (see **figure 11**) with his partner, Majd was asked later in the interview about his thoughts regarding the time he was offered prior to task performance and what he was focusing on. He said that he found the planning time really useful: "I managed to take a lot of notes which helped me remember what kind of structures would be necessary when describing the picture to my partner". He also reported that he was more focused on grammar (which he viewed as his weakest point when it comes to speaking another language) and was mostly aiming to produce accurate and well-structured sentences.

Choosing the right word/expression was another language feature that Majd mentioned he was focusing on while planning to convey the right meaning. The qualitative analysis of his notes for the first chatting session showed that Majd was taking notes in the form of short/simple sentences and that he was taking some time to edit his sentences.

Figure 11: spot the difference
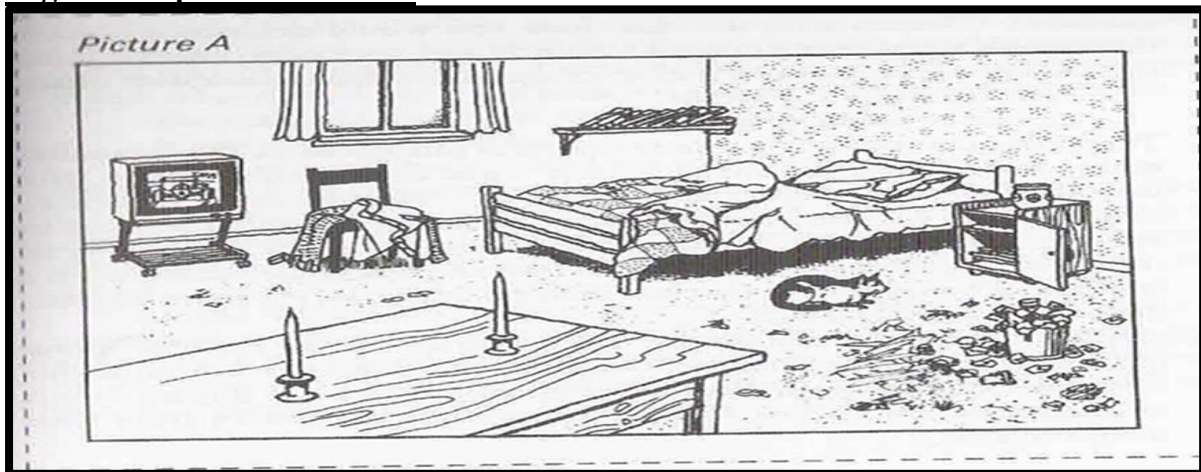
E.g., on the left down corner, there is a table with two candles on it.   (original sentence)

Down to the left corner, there is a table with two candles on it.   (revised sentence)

Majd's other planning notes for this task included the following:

On the left, there's a TV and it's on

In the middle of the room, there is a cat lying on the ground.

Next to the bed, there's a cabinet.

Majd also mentioned that he was paying more attention to his language while texting his partner, given her L1 status and the feedback that she was giving during the session (see **table 37**):

"I tried to type meaningful and accurate sentences and I was very concerned about the language that I was producing; after all, I was talking to a native English speaker, who was deliberately correcting my errors, rephrasing my sentences and telling me how a native speaker would structure a particular utterance."

E.g., Majd: My picture has a bed which is tidy.

Rosy: **In English, we can say the bed is made or not made/unmade**

Majd: ok, noted.

Rosy: OK. How about the bed?

Can you describe the cover (or duvet) that is on the bed?

Majd: it has a chess pattern and **it's unmade**.

Rosy: **We would say it has a chequered pattern**

Majd: okay, noted.

Majd emphasized that the distinctive affordances of the text chat mode, as opposed to voice chat or F2F dialogue, and the possibility to view and edit the messages before sending them to his partner aided balanced attention to different aspects of his L2 production:

"the saliency of my turns and those of my partner as they were projected on the screen, as well as the relatively short time span afforded by this mode of interaction indeed helped me attend to grammar, spelling, and/or meaning-related errors and edit my sentences accordingly. I am assuming that this would be very unlikely to happen if we were interacting F2F or even via voice chat."

Table 38: planning strategies Majd used prior and during the text chat sessions

Text-chat sessions	Spot the difference	Information-gap	Decision-making	Narrative	Decision-making	Decision-making
During the planning time	Grammar & Vocabulary to deliver meaning	Grammar & Vocabulary to deliver meaning	Content & Vocabulary to deliver meaning	Content	Content & Grammar	Content & Grammar
During the online session	Grammar Vocabulary Content Spelling	Grammar Content Spelling	Grammar, Vocabulary to deliver meaning	Content	Grammar Content	Grammar

However, as can be seen from the **table 38**, Majd's planning strategies changed once he started planning for the subsequent, more complex tasks; he seemed to focus primarily on meaning, writing short, yet incomplete sentences to highlight the main ideas. For example, below I reproduce the instructions for the first decision-making task, followed by Majd's planning notes (which seem different from the ones presented earlier on **figure 11**):

> *Lind has just returned to school and has been out of the study habit for 7 years. She has found it very difficult to get back into the habit of studying. Her time is further stretched by responsibilities of being a wife and a mother of two pre-school aged children. Based on the scenario, together with your partner <u>**identify the problem(s)**</u>. <u>**Discuss all possible solutions**</u> as well as <u>**the pros and cons (if any) for each solution**</u>.*

Majd's notes included the following:

> having responsibilities
> Manage to be a wife and a student at the same time
> Solution: taking care of the family, find a time to study
> Organise her time between being a wife and a student

During the interview, Majd reported that he was very much involved with the task itself, trying to think of as many ideas as he could to be able to discuss the problem later with his partner and find the right solutions. He commented:

> "I felt that this task was more difficult compared with the first two tasks where we already had a picture and all that is needed was to describe its content to figure out the differences between my version and that of my partner. In contrast, here we were just talking about an abstract problem and the whole conversation was mainly based on our ideas and discussion of the task's content."

Thus, according to Majd's comment, working on more cognitively demanding tasks led to a change in his planning strategies and made him pay less attention to form while planning. This appeared to be in line with Skehan & Foster's (2001) argument that "tasks which are cognitively demanding in their content are likely to draw attentional resources away from

language forms" (p.189). As shown in **table 38**, priority was given to form and word choice during the actual task performance. Perhaps the time Majd had prior to the session (which he used to address the task's goal), and the possibility to view/review the messages on the screen by scrolling the cursor backward/forward helped to free up his memory to notice and tackle communication problems caused by inaccurate language use even when working on more complex tasks.

There was, however, only one occasion where Majd chose to focus exclusively on the content of the task (narrative task) either before or during the online session. Majd explained that due to the complexity of the task and the fact that his partner was pressed for time, having a meeting immediately after the session, he found himself involved with achieving the task's goal, using the language spontaneously to manage to finish on time. During the picture-based narrative task (see **figure 12**), the pictures were divided between the participants (three or four pictures each) in the wrong order and they were asked to take turns and describe what they see in each picture, so they could decide on the right order and come up with a comprehensible story at the end.

Figure 12: narrative task (version a &b)
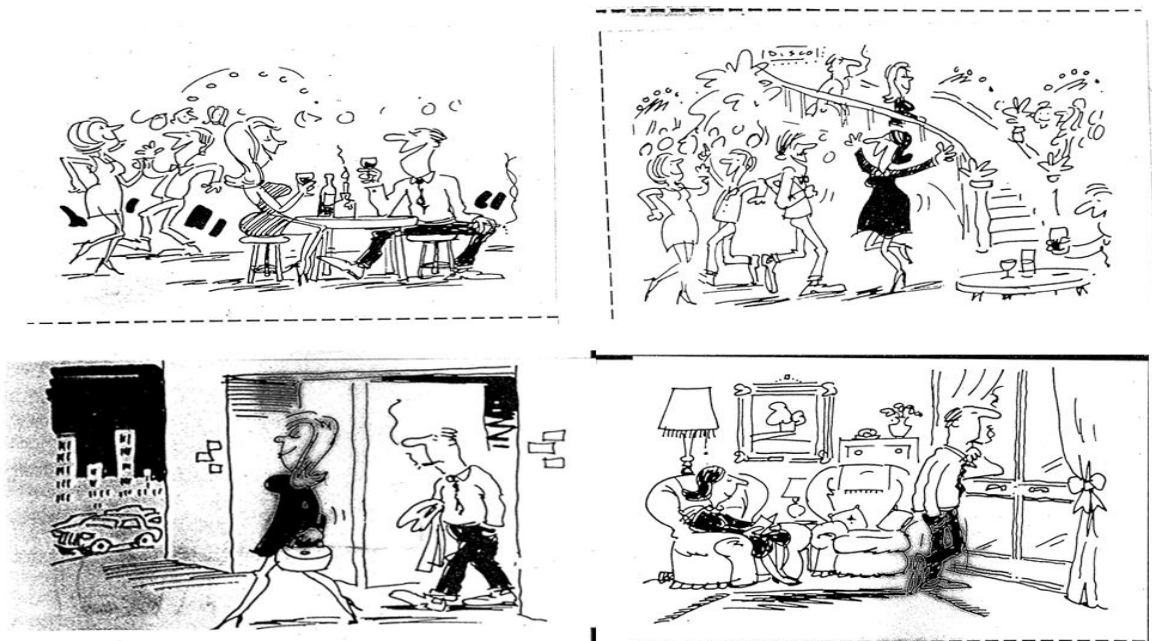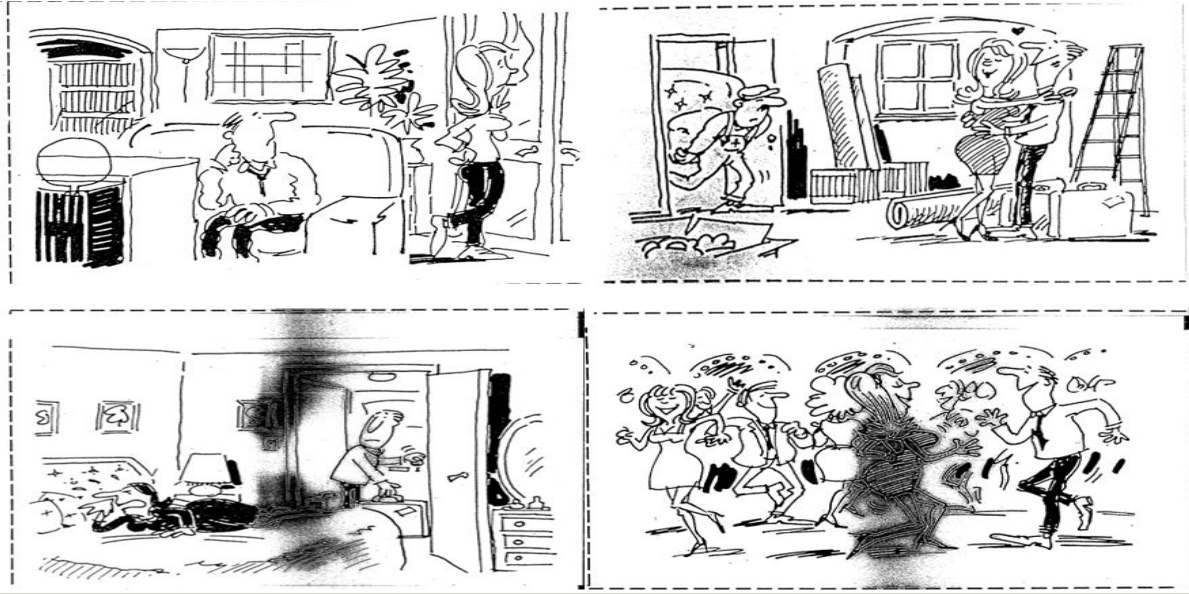
In addition, the fact that Majd did not get the chance to see all the pictures when working on this task might have made him more engaged with achieving the task goal, trying to unjumble the pictures to figure out a storyline. During the stimulated recall interview, Majd explained:

"I believe that the real challenge about this task was how to make sense of the jumbled yet incomplete set of pictures that we each had. I found it a bit hard to come up with a coherent, well-structured story based on my understanding of my version of the pictures as well as my comprehension of the content of the other pictures as described by my partner."

5.1.1.2 Case Study 2 (Lara):

Qualitative analysis of her planning notes and responses during the interviews indicated that overall, Lara's attentional resources were primarily directed towards the content (i.e., ideas/meaning) of her production while planning for the text chat sessions (see **table 39**). One exception though was found when preparing for the first task (bedroom scene, shown earlier in **figure 11**) as she reported that she was more focused on grammar, producing accurate, yet rather short sentences.

E.g., The bed is messed up

The cat is sitting beside the bed

On the right, beside the window, there's a bookshelf

Table 39: planning strategies Lara used prior and during the text chat sessions

Text chat	Spot the difference	Information -gap	Decision-making	Narrative	Decision-making	Decision-making
During the planning time	Grammar	Content	Content Spelling	Content	Content	Content
During the online session	Grammar	Grammar Spelling	Grammar Spelling	Grammar Spelling	Grammar Spelling	Grammar Spelling

Yet as the sessions were progressing, Lara seemed less concerned about the accuracy of the notes she was taking during the planning time. This was evident in her notes for the following decision-making task:

> *You and your partner have a very important and big project that is due tomorrow, and you haven't even started it yet. You plan on spending several hours doing it tonight. However, your friends suddenly ask you to go to a concert tonight. They won tickets to your favorite group and want you to go with them. You need to get the project done, but you really want to go to the concert with your friend.*
> *Together with your partner you need **to define the problem**, **explore the alternatives**, **consider the consequences**, **and identify your values**. Based on your discussion, **you should come up with your final decision**.*

Lara's notes included the following:

We will have progress working together

If we didn't do it, we'll have consequences

Either getting bad results with the project and letting the partner down or totally

failing it

Hence, Lara's decision to prioritise form over meaning as she was planning for the first session could have been influenced by the fact that she was very anxious about her language use at the beginning of this study, and that the whole online experience (practising her English online with a NS) was new to her. However, commenting on her planning notes and the strategies that she adopted prior to and during the online performance, Lara said:

> "When planning our first task, I was focusing mainly on grammar and sentence structure. However, later on I came to realise that you were asking me to use particular language forms and sentence structures, like for example relative clauses. This made the note-taking process easier for me; I mean I only took notes about the main ideas needed to achieve the task goal. However, during the sessions, I started to put these ideas in meaningful and grammatically correct sentences."

The time span afforded by the text chatting mode (though relatively short), and the possibility to see, think about, revise, and edit the messages could have also helped Lara adjust her planning strategies and make a balance between content and form in the following sessions. That is, she chose to spend the planning time to address the task's content/goal, writing down as many ideas as she could, so that she had more time to attend to the language forms and shape/structure her ideas into meaningful, accurate sentences while interacting with her partner. When asked about her decision to focus on form during the online performance, she replied:

> "I believe that note-taking had a very positive influence on me; I was mentally relieved that I got the chance to elicit my thoughts/ideas on the topic before the session started. You know, like I already had something in mind to talk about, and consequently I managed to focus on other aspects of my language (grammar, spelling and so on) when chatting with my partner."

The examples below were taken from Lara's planning notes for the third task, where she had to discuss Linda's problem, a wife and a mum of two kids, to help her get back to school after she has been out of the study habit for 7 years.

E.g., take privet lesson in some materials that she find them difficult.

She have to start with on material, easiest one

Make a study room, no disturb time

When Lara was also asked whether there were any factors affecting her planning strategies during the actual task performance, she replied:

"The fact that I was interacting with a native speaker greatly affected my choices of what to focus on (including grammar, vocabulary, etc.) …I had to think of how to express my thoughts, put them in words, and produce grammatically correct sentences, so that my partner could easily understand what I was saying."

Hence, Lara considered that attending to form during the written sessions was essential to deliver comprehensible and meaningful messages as grammatical errors might sometimes result in misunderstanding and ambiguous utterances which could eventually lead to unnecessary delay, since more elaboration would then be required to further explain what she actually meant to her partner. Thus, the L1 status of her partner seemed to have a positive impact on the language Lara was producing throughout the sessions as she had to attend to form and meaning concurrently.

5.1.1.3 Case Study 3 (Ameen)

As shown in **table 40**, Ameen also appeared to pay more attention to the content of the task (i.e., ideas) when planning for the text chat sessions. Though not reported by Ameen, the video recordings revealed that he was not completely ignoring his grammatical errors. In fact,

he was correcting these errors as soon as he noticed them. For example, Ameen first wrote _she is too young and have_ but then revised it to _she is too young and she has_. He also wrote _this is who should have a heart_ and revised it to _this is who should have the heart_.

Table 40: planning strategies Ameen used prior and during the text chat sessions

Text chat	Spot the difference	Information-gap	Decision-making	Narrative	Decision-making	Decision-making
During the planning time	Vocabulary	Vocabulary Spelling	Content	Vocabulary	Content	Content
During the online session	Grammar	Grammar	Grammar	Content	Grammar Content Spelling	Content Spelling

Ameen reported that while planning, he was also attending to vocabulary (word choice) to express his thoughts, particularly with picture-based tasks (i.e., narrative, spot the difference, and information-gap) as he had to be precise in describing what he could see in the pictures to his partner. When asked about the aspect(s) of his language which he was focusing on while planning for the first task, Ameen responded:

> "I was mainly focusing on the appropriate words/vocabularies to describe what I had in my pictures. I was also thinking about the right expressions that should be used to connect sentences together like for example, on the left side, in the corner, and so on."

It is worth mentioning, however, that Ameen spent only 6 minutes of the planning time trying to type some notes for the first task and eventually he just wrote: _there is a table, 2 candle, wastbasket_. This happened very often as Ameen was planning for the other sessions; he seemed very hesitant when it came to note-taking. It had been assumed that the lack of note-taking skills could have made it difficult for some learners to generate more notes on the assigned task. Although the learners were given instructions at the very beginning of the study on how to take notes, they did not seem to be very familiar with this kind of skill; as shown in the

examples provided earlier for each learner, most of the notes were in the form of full sentences which they used later during the online sessions.

During the interview, Ameen was asked for a justification of what was going on and why he was reluctant to take notes. He replied by saying that he did not see the point of putting his thoughts into words before the session and that he would better organise the ideas that he had about a task in his mind. Hence, perhaps at this point, Ameen was still not considering the time that he was given to plan his online performance necessary prior to the text chat sessions since the mode itself could afford some sort of online planning (time to think of what to say) during the actual performance. Working on different and more complex task types (i.e., decision-making tasks), though Ameen was not yet fully benefiting from the planning time (i.e., only spending four/five minutes taking notes), his notes became more like short sentences. Commenting on his note-taking behaviour at this stage of the study, Ameen said: "I felt like somehow it would be useful to take notes here as the task itself depended more on my ideas". Another sign of progress in terms of note taking was acknowledged by Ameen as he was preparing for the final text chat session; where the participants had to decide on whether to stay at home and get their project done for next day submission or go to a concert with their friends. Ameen was noticeably faster and managed to take more notes during the planning time compared with the previous sessions, spending five minutes on the task. He wrote the following:

Try to specify the most important thing in the project

if you go, you can stay up all the night doing your project

if your partner doesn't want to go to the concert, he can do the half of it and when you come back, you can complete the rest

maybe I can go with them and ask them to help with my project

ask my partner what does he think

Ameen reported that as he could relate to the task's scenario, he had many ideas in his mind, and therefore, he felt comfortable typing more notes. Thus, it seems that task type and content also influenced how/whether Ameen perceived the benefits of the planning time.

The planning strategies Ameen used during the online sessions, however, were a bit more varied; his attention was directed towards grammar and content. Attending to spelling was also a recognisable trait while Ameen was interacting with his partner; he frequently attempted to self-correct his spelling errors and/or use the auto-correction feature when he could not figure the right spelling out. Ameen was also amongst the participants who prioritised form over meaning in most of the sessions because of the high proficiency level of his partner, commenting:

> "My partner is an English teacher and his English is flawless, so whenever possible I was focusing more on grammar than any other aspects of my language. Hence, the relatively long time that I was taking to type my messages was basically because I wanted to make less errors."

5.1.1.4 Case study 4 (Rima)

Predictably, Rima, who was in the control group (i.e., did not prepare for the tasks beforehand) focused primarily on her word choice to deliver meaning regardless of the complexity and type of the task (see **table 41**). Following the first session, Rima commented:

> "The most challenging part for me was finding the appropriate word(s) to describe something because I always have the feeling that I don't have a vast vocabulary bank and that I should be learning more."

Table 41: planning strategies Rima used during the text chat sessions

Text chat	Spot the difference	Information-gap	Decision-making	Narrative	Decision-making	Decision-making
During the online session	Vocabulary	Vocabulary	Vocabulary	Vocabulary Spelling Content	Vocabulary	Vocabulary

Hence, given that Rima was not asked to attend to her language use or offered any guided instructions to integrate particular linguistic forms, she was predominantly involved with achieving the task's goal and thus not finding the right words to express her thoughts to her partner was making her very nervous during the sessions. As illustrated in the example below, Rima was less concerned about the accuracy of her production than her word choice:

> E.g., Ok in my opinion, he is a doctor and actually I think other patient deserve it more
>
> than him. He might has acquaintances in his field, so he can get another heart.

Thus, it seems that giving learners time and guided instructions before the text chat sessions to prepare for a particular task might help them address multiple aspects of their production, balance between form and content, and subsequently facilitate their online interaction with more proficient L2 speakers. That is, as a result of the time/instructions, learners would have the opportunity to address the task's content, taking notes of the main ideas that they need to achieve the task's goal, while at the same time freeing their attention to focus more on the linguistic aspects of their production later in the online session.

5.1.2. Planning strategies for different task types before and during the voice chat sessions

Similar results to those found during the written phase were also noticed as the learners were preparing for their oral interaction. Overall, all the three learners who received guided planning instructions prioritised content and word choice over form regardless of the task type when preparing for the online sessions. However, Majd and Lara managed to pay more

attention to form during the actual task performance. All the planning choices made per case study before and throughout the voice chat sessions will be fully presented below.

5.1.2.1 Case study 1 (Majd)

Majd, who was deliberately focusing on grammar while preparing for the written tasks, seemed to follow a new strategy of note taking for the audio sessions (see **table 42**). In almost all the tasks, he reported that he was primarily focusing on the content of the task, i.e., all the notes were targeting the main ideas that he needed for the oral discussion with his partner.

Table 42: Majd's planning strategy prior and during the voice chat sessions

Voice chat	Spot the difference	Decision-making	Narrative	Decision-making	Decision-making	Decision-making
During the planning time	Vocabulary	Content	Content Pronunciation	Content	Content Grammar	Content
During the online session	Vocabulary Content Grammar	Content Vocabulary Grammar	Content Grammar Pronunciation	Content Grammar	Content Grammar	Content Grammar

This shift in Majd's focus was understandable given the mode of task delivery (oral not written). To put it another way, unlike the written chat, the conversation was more instant, and the learners did not have much time to think of new ideas during the oral interaction which was why they preferred to do that during the planning time. This was supposed to relieve some of the pressure that the learners might have encountered during the sessions owing to the task demands, learners trying to attend to multiple aspects of their production, and/or the instantaneous nature of the audio interaction which somehow resembles F2F conversation. For example, Majd seemed very anxious when he was preparing for the second task which was discussing a very common topic (smoking):

*Your friends (including someone you have a crush on) start smoking. What would you do? Together with your partner you need to: **discuss the reasons that you think might make them take that decision**, **consider the consequences**, **and identify your values**. Based on your discussion, **you should also think of ways to convince your friends to give up on that bad habit.***

Preparing for this task, Majd ended up only writing the following notes during the planning time:

Pressure, stress, life crises

Provide some sort of **calm-ness**, it **ease them up**

Hurting other people, themselves **on** the long run

Alternative: e-cigarettes

Later in the interview, he responded as follows when asked about what he was thinking of and what aspects of his language he was focusing on during the planning time:

"I was completely focusing on the ideas that are related to smoking. I put myself in the smoker's place and tried to imagine what would make me take that decision and smoke. However, due to the lack of knowledge that I had about this topic, I could not take a lot of notes and therefore, I was feeling very uncomfortable as the session began."

It is worth mentioning, however, that as Majd was discussing the topic with his partner, he mentioned that his dad was diagnosed with lung cancer because he was a heavy smoker and eventually, he lost him to that. This could have stirred up a lot of bad memories for him and might have negatively affected his ability to recall several ideas on the topic. Thus, it seemed that the more notes the learners managed to take prior to the task, the more relaxed they felt during the session.

It is also worth stating that Majd was showing progress with regards to his notetaking skills during the voice chat as compared to the text-based equivalent; his notes were more in the form of key words that represented main ideas rather than full sentences.

E.g1., <u>**Notes during text chatting:**</u>

first patient: yes, cuz he's at the height of his career yet no cuz he's a surgeon and he probably has more access to med info than others and he can figure it out

E.g2., <u>**Notes during voice chatting:**</u>

rough childhood, abusive parents, wrong parenthood-ing, mentally ill and doesn't realise it. Cause depression, physical harm maybe

Majd, however, was not entirely overlooking the accuracy of his production. He mentioned that he was worried about his ability to produce meaningful and grammatically correct sentences, especially with discussion-based tasks:

> "Working on the tasks via voice chat was different; there was no time for me to think about and edit my sentences like I did in the written chat. Therefore, I was really concerned about my language and I was hoping not to make major grammatical errors which could in turn affect the clarity of my messages".

Though the learners were interacting orally with their partners and did not have much time to think of their production compared with the written sessions, as shown in **table 42** above, Majd reported that he was attending to grammar as well as content and vocabulary during the audio sessions. His ability to attend to multiple facets of their L2 performance was improving throughout the audio sessions. For instance, during the third audio session, Majd noticed that he was progressing in terms of the quantity and quality of his production being more complex, accurate, and fluent. He commented:

> "Concentrating on grammar and meaning simultaneously was the toughest part for me in the previous oral sessions; I was literally overwhelmed with the instant nature of interaction, the task itself, and my partner being a native English speaker. However, in

this session, I started to feel less anxious, yet more capable of balancing my attentional resources to focus on different features of my production."

5.1.2.2 Case study 2 (Lara)

Qualitative analysis of the planning strategies Lara used while preparing for the voice chat sessions were consistent with the ones she adopted for the written chat; she reported that she was attending to the content of the tasks as well as her vocabulary choice (see **table 43**). Hence, it appears that no difference was found in the planning choices she made due to the change in the mode of task delivery.

Table 43: Lara's planning strategy prior and during the voice chat sessions

Voice chat	Spot the difference	Decision-making	Narrative	Decision-making	Decision-making	Decision-making
During the planning time	Vocabulary Content	Content Vocabulary	Content Vocabulary	Content Vocabulary	Content Vocabulary	Content Vocabulary
During the online session	Content Vocabulary	Content vocabulary	Vocabulary Content grammar	Content Vocabulary Grammar	Content Vocabulary Grammar	Content Vocabulary Grammar

Lara, who failed to produce error-free structures while taking notes, was mostly focused on her ideas. She was also very particular about her word choice as she kept editing her notes. This was noticeable as she was planning for the 'moving abroad' task where she had to think of the pros/cons of moving to another country and decide on whether to take this opportunity that she had been offered, leaving everything behind and starting over a new life.

    E.g., To have a better job, better financial future (original version)

        To have better future career. To improve their financial situation (revised version)

Similar to Majd, Lara also reported that she was trying to attend to grammar as well as content and vocabulary during the audio interaction. As the sessions were progressing, Lara was getting

more motivated and was showing greater capacity to balance attention on multiple facets of her L2 performance. Following the second session, for example, she recognised some grammatical errors while listening to the recordings of the previous sessions, and therefore she began to pay more attention to form. Lara said:

> "I was hearing myself making lots of errors related to tense, sub-verb agreement, etc., This was due to the lack of practice, I guess, so I will try to be more attentive to grammar in the coming sessions."

Thus, she modified her planning choices during the online performance as she noticed the gaps she had in her L2 due to the scarcity of L2 speaking opportunities in her context. Lara also traced this back to the mode of task delivery which was more rapid than text-based chat; hence the conversation was apparently more spontaneous and natural.

It is worth mentioning that Lara appreciated the planning time she had been offered prior to each session as it was helping her highlight the main points that were yet to be discussed with her NS partner. At the end of the study, she reported:

> "I loved the part where we got the chance to take notes. I found it very beneficial to brainstorm thoughts/ideas on the task's content beforehand, and it actually helped me direct my attentional resources afterwards to the gaps that I noticed in my L2."

5.1.2.3 Case study 3 (Ameen)

Ameen also reported that he was primarily focusing on addressing the task's content when planning for the oral sessions (ideas/word choice) (see **table 44**). He said:

> "I was quite worried about my ability to keep talking on a particular topic for like twenty or thirty minutes and I've always had the feeling that I do not have much to say to keep the conversation going, which was why maybe I was focusing more on the content of the task during the planning time".

Hence, based on Ameen's comment, the mode of interaction seemed to have some impact on the aspect(s) of language that he chose to attend to during the planning time. These were not found to be very different though when compared with his planning strategies for the text-based chat sessions.

Not much change was recognised in terms of the quality and the quantity of the notes Ameen was taking during the planning time, i.e., by the end of the study, Ameen's notes were still more like complete, somehow long, sentences; most of which were used afterwards in the online sessions without any further elaboration. For example, in the last audio Skype session the interlocutors had to discuss the following task:

> *Your friends are bullying someone at university/work. You used to be friends with the person that is being bullied. What would you do? Together with your partner you need to discuss,*
>
> ➢ *What bullying means and what kind of actions does it involve?*
> ➢ *Why do you think some people like to bully others?*
> ➢ *Any short term/long term effects you think this action could have on the person who is being bullied*
> ➢ *Would you try to convince your friends to stop that action? How?*
> ➢ *What would you do if they refused to listen to you?*
> ➢ *How your personal values and beliefs affect your decision*

During the preparation time that he was given before the session commenced, Ameen wrote:

bullying means hate, insult the person infront of you or caused them harm

To show that they are stronger or who has the power to do anything they want

Depression, hate themselves, lack of self-confidence

Voice chat	Spot the difference	Decision-making	Narrative	Decision-making	Decision-making	Decision-making
During the planning time	Vocabulary	Content	Vocabulary Content	Content	Content	Vocabulary Content Grammar
During the online sessions	Vocabulary	Vocabulary Grammar	Content	Vocabulary	Content	Content

Ameen also did not report any change in his task approach during the online interaction; he mentioned that his main focus while chatting with his partner was on delivering meaning and achieving the task goal. Possibly the fact that his partner was not giving him any feedback on his errors made him less concerned about the quality of his production. Additionally, the pressure caused by the oral interaction (i.e., the lack of thinking time) since the conversation was more instantaneous than the written chat, as well as the variety/complexity of the tasks, could all help account for why Ameen prioritised one characteristic of the language over the others.

5.1.2.4 Case Study 4 (Rima):

Rima, who was in the control group, was also completely focusing on vocabulary during the audio sessions (see **table 45**). Discussing the bullying task with her partner, Rima wrote:

> **E.g.,** we have to explain the meaning of the word, so to start with that I think it's some actions and some maybe terms or vocabs that you use to hurt the one that you maybe hate or…

During the interview, she commented:

> "Finding the appropriate words to talk about your feelings in another language is a very challenging part for me, all the other things like grammar for example could come later. I mean without a sufficient amount of lexical items, you could go nowhere with your

conversation. Maybe I made too many errors while speaking, yet the good thing was that my partner was still able to comprehend my messages."

Table 45: **Rima's planning strategy during the voice chat sessions**

Voice chat	Spot the difference	Decision-making	Narrative	Decision-making	Decision-making	Decision-making
During the online session	Vocabulary pronunciation	Vocabulary pronunciation	Vocabulary pronunciation	Vocabulary	Vocabulary	Vocabulary

Thus, it seems that Rima's choices of what to focus on while interacting with her partner were primarily influenced by her desire to achieve the task's goal and practise her English by producing more appropriate language throughout the online interaction. The fact that Rima did not receive any guided instructions with regard to her language use could have made her less attentive to the formal aspects of her production as well.

5.1.2.5 Summary

All three learners in the planning group confirmed that their focus during the preparation time remained largely unchanged as the tasks increased in complexity, in terms of type, content, and mode. Hence, these findings revealed that the learners originally concentrated on meaning and retained their attention on meaning as they planned for more complex tasks over time. Occasionally, the learners also attended to other aspects of their language such as grammar, vocabulary, and pronunciation. This contradicts the findings of Ortega (2005) and Thompson (2014); in these studies, the learners who were provided with grammar guidance focused primarily on form and tried to use the targeted structures while planning for the narrative tasks. This could be traced back to different factors, such as the mode of task delivery (written and oral via SCMC, not F2F as in the previous studies). There was also the variety of task type the learners worked on as compared with previous research where the learners mainly performed narrative tasks (people going to buy a pet, a car, a toy and so on). That is to say, the complexity

of their tasks was mainly in the number and variety of the targeted structures the learners should employ during the narration. This could have made it easier for the learners to take notes and could have directed the learners to pay more attention to form. Finally, there was the fact that learners in this study were working on dialogic rather than monologic tasks. In other words, achieving the task's goal was based more on the interaction between two participants for a duration of twenty to thirty minutes rather than narrating in one minute. Thus, language production throughout the online sessions was more spontaneous compared with previous research and resembled real-life conversation.

However, learners' attentional resources varied across the online sessions: some of them made a greater use of the planning time, showing more control over their production and balancing their attention between form, meaning, and word choice, whereas others were more focused on achieving the task goal. It is worth noting that task type was not reported as an influencing factor for what aspects of production learners chose to attend to before and/or during the voice chat sessions, yet it seems that there were other factors that could have contributed to this decision (e.g., guided planning, mode of task delivery, feedback received during the session, motivation and anxiety level). These will be fully discussed later in this chapter.

Having addressed learners' planning strategies, the following section will elaborate on the factors that impacted the four learners' L2 performance during the text-based and voice chat sessions, including their production of more/less complex, accurate, and fluent output.

5.2 RQ4: What factors, if any, impacted learners' L2 development across different modes of online interaction (text vs. voice chat)?

Logs consisting of recorded online sessions were manually transcribed as Word files and qualitatively analysed to detect any changes in learners' production throughout the online interaction. It should be noted though that before concentrating on the qualitative data, some

quantitative figures will be included in this part of the results to set the scene and gauge the progress (if any) achieved by each of the four case studies with regard to the complexity, accuracy, and fluency of their production as the online sessions were progressing. Due to the nature of written chat (which is based on turn-taking between two people), the lack of access to keystroke logging software (to capture and monitor learners' actions on the keyboard while they were typing), and the space/time limit of the current research, only a few CAF measures were used to qualitatively assess learners' production throughout the online sessions. These measures involved learners' integration of relative clauses, the number of errors they made, as well as the number of pauses identified per session. The results are presented accordingly in the following sections divided per mode of interaction.

5.2.1 Text-based chat

An analysis of the aggregated data (chat logs and stimulated recall interviews) showed a number of differences amongst the SLs in terms of the quality of the language they produced while interacting with more proficient L2 speakers via means of text-based chat. The learners also exhibited different behaviours in response to the L2 practice opportunities they were offered in the current study. Apart from the different treatments they received (i.e., planning group VS control group), a number of other factors were found to impact their L2 performance/development including the feedback they received during the sessions (if any), the task type/content, the rapport with their partners, and their motivation/anxiety levels.

5.2.1.1 <u>Complexity measures</u>

As expected, the three learners who received guided planning treatment tended to employ more relative clauses while interacting online with their partners, compared with Rima who was in the control group. There was, however, a lack of consistency in the number of

relative clauses produced by each learner throughout the sessions. **Table 46** below shows that Majd generated more relative clauses than the other two learners who also worked under the planning condition, with an overall number of 24 instances for Majd, 12 for Lara, and 15 for Ameen as against 5 instances for Rima (who worked in the control group).

Table 46: Complexity during the text chat sessions

Tasks	Groups	Pairs	Number of Relative Clauses
Spot the difference	TG	Majd & Rosy	3
		Lara & Kate	2
		Ameen & Rex	1
	CG	Rima & Rana	0
Information-Gap	TG	Majd & Rosy	5
		Lara & Kate	2
		Ameen & Rex	1
	CG	Rima & Rana	0
Decision-making	TG	Majd & Rosy	4
		Lara & Kate	2
		Ameen & Rex	1
	CG	Rima & Rana	2
Narrative	TG	Majd & Rosy	7
		Lara & Kate	3
		Ameen & Rex	4
	CG	Rima & Rana	1
Decision-making	TG	Majd & Rosy	4
		Lara & Kate	3
		Ameen & Rex	6
	CG	Rima & Rana	2
Decision making	TG	Majd & Rosy	1
		Lara & Kate	0
		Ameen & Rex	2
	CG	Rima & Rana	0

Qualitative analysis of the chat logs revealed that Majd's partner, Rosy, was frequently giving him feedback during the session, and pushing him to rephrase his utterances using these forms. Besides, Majd often reported that he was intentionally copying his partner's style, using similar structures, so that his L2 production would sound more native-like. He also mentioned that he was fully aware of the corrections given by his partner, especially those related to the

integration of relative clauses. This in turn could have positively influenced the number of relative clauses he produced per task.

> **E.g.,** Majd: On the second shelf, there are a CD collection
>
> Rosy: because 'collection' is singular, what would you need to say instead?
>
> Majd: collections?
>
> No no
>
> Wait
>
> Rosy: no, collection is singular, so you need to change the verb
>
> Majd: there is a CD collection
>
> Rosy: yes!
>
> Great
>
> Please continue
>
> sorry, **you can also say 'there is a CD collection which is on the second shelf'**
>
> Majd: [on the] [22]and there is a third shelf, **which has a sound system on it**

As shown in the example above, Majd started the last turn with *on the* but then changed his mind; he revised the structure of the sentence and used a relative clause. When asked to comment on this during the interview, he said:

> "I was typing something but then I noticed what my partner wrote in her last message and how she rephrased my previous sentence, so I just tried to copy her and produce a sentence of a similar structure".

Additionally, one could presume that there was a relationship between task type and learners' production of these forms during the written sessions as more relative clauses were employed by the learners while working on the fourth and fifth tasks (narrative and decision-making). It

---

[22] Texts located in brackets indicated text that was produced but then deleted before sending.

is worth stating, however, that unlike previous research (e.g., Ortega, 2008; Thompson, 2014) which mainly used narrative tasks that have a similar story pattern (i.e., a family going to buy a car, a pet, and so on), the variety of task type or task content *per se* in the current study meant that not all the tasks would inevitably necessitate a natural and an equal integration of relative clauses into learners' L2 production. This could explain the fewer relative clauses employed by the learners in the last session (which was also a decision-making task).

The chat logs also revealed that not all the NSs and Syrian teachers were following the given instructions and frequently encouraging their partners to utilise relative clauses during the online sessions. Rosy, for example, who was trying to trigger Majd's attention to use these forms, mentioned that sometimes she was very hesitant and cautious when explicitly recommending their use:

> "I felt sometimes they were speaking very well using relatively complex language. I did not see that they were necessarily speaking badly sometimes, so having to try to come up with things on the spot to reformulate, to make it more complex sometimes felt a bit artificial. It is like I am doing this unnecessarily. But if that is their focus, if that is what they want to improve the most, then I think I could have done more."

Others chose either to ignore the instructions altogether, or to be implicit and produce these forms while interacting with their partners, hoping that the latter would eventually notice the way they structured their sentences and try to copy them. Unfortunately, most of these indirect attempts either went unnoticed by the learners, or that the learners noticed the corrections but chose not to copy them themselves.

In fact, the guided instructions, provided to the treatment group before the beginning of each session, were sometimes the main reason behind learners' implementation of these forms.

Having read the language related instructions which urged the learners to integrate particular language forms, Lara decided to adopt a new strategy to approach the requirements and the cognitive demands of the tasks. Thus, rather than focusing on multiple aspects of her L2 production at once, she ended up taking more notes on the content of the task during the planning time so that she could concentrate more on her language use and address the targeted structures afterwards. During the stimulated recall interview that followed the second chatting session, Lara commented:

> "I realised that you were asking me to focus on particular language forms and sentence structures, like for example using relative clauses. Hence, during the session, I remember that I made a few attempts to use these forms".
>
> **E.g.,** Remember the happy couple **who got new furniture**?
>
> **E.g.,** The 17 years old waitress **who has a disabled brother**

In addition, Ameen, who did not receive any feedback from his partner, sometimes edited his sentences to add a relative clause. Commenting on these revisions, Ameen said: "In the instructions, you mentioned that bit about utilizing particular structures while talking to our partners, so whenever possible, I tried to edit my sentences for these structures".

> **E.g.,** the first patient **who is a brain surgeon**, he can help the other patients and maybe he can help them to live longer.
>
> **E.g.,** but I guess he will find someone else **who can take care of him**

In sum, it seems that the guided planning, task type/content and the treatment received by the learners in the planning group during the text chat sessions (i.e., feedback) had some positive influence on complexity compared with Rima who worked in the control group. These findings align with the quantitative results (discussed earlier) that showed a greater implementation of complex structures by the treatment group compared with the control group during the post-

/delayed written tests. Perhaps learners' practising to use relative clauses by performing tasks that varied in their type and complexity led to the automatisation of the targeted language throughout the tests. The findings also comply with those of previous research which also reported gains in syntactic complexity due to guided planning (e.g., Kawauchi, 2005; Mochizuki & Ortega, 2008; Sangarun, 2005; Thompson, 2014). Thompson (2014) concluded that "strategic planning can have a more positive impact on complexity if learners' attention is directed towards a specific linguistic form and specific measures are used to track learners' production of it" (p.245). Unfortunately, a full comparison between the findings of the current research and those of previous research was not possible due to the different treatment the learners received as concerned mode of interaction (i.e., SCMC vs F2F), task delivery (dialogic rather than monologic), variety of task types, as well as the fact that the instructions delivered to the NSs/STs were conformed to rather unevenly.

Having reviewed the complexity results of this study, the next section will address the accuracy findings of learners' production in the text chat sessions.

5.2.1.2 <u>Accuracy measures</u>

All the errors made by the learners were counted and logged, whether these were grammatical, spelling, or word/content related errors. The data showed inconsistencies as to the progress achieved by the treatment group throughout the sessions (see **table 47**). To put it another way, there was no noticeable decrease in the number of errors the three learners were making while they were interacting with more proficient L2 speakers. On the contrary, compared with the first session, more errors were sometimes identified as the online sessions proceeded. Unsurprisingly, Rima was also making more errors as the sessions proceeded.

Table 47: Accuracy in text chat sessions

Tasks	Groups	Pairs	Number of errors
Spot the difference	TG	Majd & Rosy	11
		Lara & Kate	20
		Ameen & Rex	8
	CG	Rima & Rana	18
Information-gap	TG	Majd & Rosy	10
		Lara & Kate	35
		Ameen & Rex	12
	CG	Rima & Rana	38
Decision-making	TG	Majd & Rosy	14
		Lara & Kate	16
		Ameen& Rex	17
	CG	Rima & Rana	37
narrative	TG	Majd & Rosy	11
		Lara & Kate	27
		Ameen & Rex	17
	CG	Rima & Rana	22
Decision-making	TG	Majd & Rosy	11
		Lara & Kate	16
		Ameen & Rex	18
	CG	Rima & Rana	42
Decision making	TG	Majd & Rosy	7
		Lara & Kate	15
		Ameen & Rex	27
	CG	Rima & Rana	17

It was not until the final text chat session when all the learners (apart from Ameen) produced fewer errors than the previous sessions. This lack of consistency in the number of errors the learners produced per session could be traced back to different factors. NSs'/STs' reluctance to frequently provide feedback on their partners' errors could have made the learners less attentive to the accuracy of their production and resulted in a higher number of errors. When asked about his focus while interacting with his partner in the first two sessions, Ameen reported that he was focusing primarily on form. He was not taking risks and produced mostly short/simple sentences, hence only eight errors were captured in the chat log of the first session.

Yet in the following sessions, Ameen, who had hardly received any feedback on his language production, shifted his focus and began to prioritise meaning over form, generating longer yet less accurate turns. Hence, possibly the lack of feedback from his partner could have had a detrimental influence on the accuracy of Ameen's L2 performance. The example below, where the interlocutors were discussing Linda's problem to help her get back into the study habit, demonstrated how Rami was not commenting on Ameen's language or drawing his attention to the errors he was frequently making during the sessions.

E.g., Ameen: I think it is a good idea to send her children to kindergarten when she thinks that she doesn't have time for them but she doesn't have to depend on kindergarten and send them to always because the children need their mother's care

Rami: Ok good ideas. What if she can't afford it?

On the other hand, Majd, who did receive corrections on his errors, reported that he was deliberately attending to form while chatting with his partner, firstly because he already knew he was not very good at grammar, and secondly because his partner was a NS and was correcting him very often. In the examples below, Majd and Rosy were discussing the circumstances of six patients who need a heart transplant. There is only one heart at this time, so they had to agree on who should receive the heart.

E.g., Majd: Yet, I think that he can find another heart due to him being in the medical field so he must be able to access and reach various people and information about another donor.

Rosy: that's an interesting point

Since we don't know for sure, it's appropriate to say 'I think that he COULD find another heart'

Majd: ok

E.g., Majd: Well, I think the teacher should receive it because first he has two

children that he has to take care of and he is basically teaching generations

so he's a treasure for sure

Rosy: I agree with you there

I don't think we would say 'he's a treasure'

we could say 'he's valuable to society/he makes a significant contribution

to society'

Majd: yes

As shown in the two excerpts above, Rosy's feedback was not only covering the grammatical errors that Majd was making, but also other types of errors including inappropriate word choice (i.e., words that were inaccurately used in the given context).

Task type/content seemed to influence the quality of learners' production as well, i.e., picture-based tasks that required the learners to frequently use specific structures and non-salient forms to describe their pictures (e.g., prepositions of place, articles, third-s singular), led to a higher number of errors than decision-making tasks (see **figure 13**). The learners reported that the latter, though sometimes more cognitively demanding, allowed more flexible discussions between the interlocutors compared with the other task types.

Figure 13: information-gap task

E.g., Lara: but there**'s** 3 shelves above it

let's start with the first one **from the above**

Kate: okay the top one?

Lara: there`re many books **on the both sides of the shelf** but we can see a clock

 in the middle

the second one same thing but there`s a CD in the middle

the third one which is **the lower one**

Kate: so there are books on the top two shelves?

We see how Kate ignored Lara's grammatical errors (singular/plural verb, article) and only provided implicit feedback on her partner's word choice. It is worth mentioning, however, that most of the treatment group's grammatical errors while performing the tasks were associated with articles, prepositions, subject-verb agreement, and singular vs. plural. In other words, these errors were not affecting the clarity of their messages, and therefore NSs/STs may have therefore judged it pointless to give feedback on these errors and interrupt the flow of interaction.

Rima, who did not receive guided instructions and was not asked to attend to her language, unsurprisingly produced more errors compared with the treatment group. During the

229

interviews, she said: "focusing on grammar was not a priority for me, rather finding appropriate vocabulary to express my thoughts was my primary concern across the chatting sessions".

E.g., Rima: Yes, she is poor girl like she haven't get enough from her life yet

Rana: Yeah

Rima: One the other hand, I think she might find a charity which works for children to help her

Hence, one could conclude that over time the online treatment appeared to reduce the number of errors the learners in the treatment group produced during the text chat, which in turn resulted in greater gains in accuracy throughout the immediate and delayed post-tests as shown by the quantitative data in the previous chapter. The qualitative analyses of learners' production during the written chat shed light on what reasons could have affected learners' production of more or fewer errors. As discussed above, these involved the frequency and the type of feedback provided to learners during the sessions, if any. Task type was an additional factor that seemed to influence the accuracy of learners' production as some tasks triggered the learners to produce more elaborated and complex language than others (e.g., decision-making vs. spot the difference).

The following section will, therefore, discuss the fluency of learners' production during the text-based chat.

5.2.1.3 <u>Fluency measure</u>

As stated earlier, fluency could not be properly measured during the written online sessions due to the lack of access to keystroke logging software. Hence, learners' pauses (filled/unfilled) throughout the written sessions were counted and logged for further analysis. This measure by no means was intended to gauge the extent to which the output produced by

the learners could be classified as fluent or otherwise. In fact, the qualitative analysis for this measure only aimed to understand why the learners were pausing, i.e., what was going on in their mind every time they paused during the sessions.

Table 48: Fluency in text chat sessions

Tasks	Groups	Pairs	Number of Pauses
Spot the difference	TG	Majd & Rosy	14
		Lara & Kate	20
		Ameen & Rex	16
	CG	Rima & Rana	7
Information-gap	TG	Majd & Rosy	5
		Lara & Kate	8
		Ameen & Rex	10
	CG	Rima & Rana	10
Decision-making	TG	Majd & Rosy	13
		Lara & Kate	31
		Ameen & Rex	23
	CG	Rima & Rana	12
narrative	TG	Majd & Rosy	11
		Lara & Kate	22
		Ameen & Rex	39
	CG	Rima & Rana	13
Decision-making	TG	Majd & Rosy	5
		Lara & Kate	28
		Ameen & Rex	26
	CG	Rima & Rana	13
Decision making	TG	Majd & Rosy	6
		Lara & Kate	20
		Ameen & Rex	22
	CG	Rima & Rana	12

As shown in **table 48**, the number of pauses generated per session varied amongst the learners. Some of the learners paused more as the tasks increased in their complexity, i.e., decision-making and narrative tasks required more production than spot the difference and information-gap tasks; therefore, more pauses were detected in the former. Other learners, however, were found to pause less as the online treatment proceeded. Accordingly, the learners were asked during the interviews to recall their thoughts on all the identified pauses and hesitations in their online performance; different explanations were provided. Thinking about word

choice/spelling, sentence structure, and other linguistic forms (e.g., tense, articles) while describing the task pictures were amongst the reasons that frequently caused hesitation; that is, the learners sometimes paused to read and edit their sentences. Majd, for example, appeared to be hesitant when typing the sentences below.

> **E.g.,** Majd: in the down left corner, there's a table with two candles on it and it's (3
>
> seconds pause) not lit
>
> Rosy: you mean, 'in the bottom left corner'?
>
> Yes, I have a table with two candles on it
>
> but they are not lit
>
> Majd: yes exactly
>
> Rosy: You could also say, 'on the table there are two candles which are not lit'
>
> Majd: In the bottom right corner, there's a garbage can (2 seconds pause) and it *is*
>
> *full and it* has rubbish (1 second pause) all around it (10 seconds Pause) (original)
>
> In the bottom right corner, there's a garbage can, which is full and there's rubbish
>
> all around it.   (revised)

When asked about the pause that was captured when writing the first message in the example above, Majd explained:

> "I was almost confident that *down* was not really the right word to use here so I paused
>
> because I was trying to think of a more appropriate alternative, but then to avoid any
>
> further delay in sending my message, I decided to leave it as it is and wait for my
>
> partner's feedback."

As he was proceeding with his description, more pauses were identified. Majd mentioned that the first pause was because he was a bit uncertain about the spelling of the word *can* whereas the second pause was when he reread the whole sentence and thought that some details were

still missing, so he went back and wrote *is full and it*. Commenting on the very long pause that

he made at the end, however, Majd said:

> "Just before sending what I wrote to my partner, I realised that she rephrased my
>
> previous message using a relative clause and so it took me a few more seconds to figure
>
> out how to revise and structure my sentence accordingly."

Hence, the fact that Majd's partner was constantly giving feedback on these structures made

him more preoccupied with these forms, yet less fluent sometimes. Fewer pauses were

captured, however, as the online sessions proceeded (see **table 48**). Perhaps Majd's continual

attempts to integrate relative clauses into his L2 production boosted the retrieval of these

structures. According to Gilabert (2007b), this in turn could have led to "more efficient

message planning and faster lexical access and selection", which consequently benefitted

fluency (p.64). Discussing the circumstances of six patients to decide on who would be eligible

to receive a heart transplant, the interlocutors wrote:

> **E.g.,** Rosy: what about the 17 year-old who cares for her brother?
>
> I think it's admirable that she cares for her disabled brother
>
> Majd: for me I think she's the one who should receive it because she's taking care
>
> of her disable brother, yet at the same time I feel like if something happened
>
> to her maybe a caring center or an establishment for disable people could
>
> take him in
>
> Rosy: So, having the heart could help both her and her brother
>
> Majd: now as for the female AIDS researcher, maybe we should choose her because
>
> she works to find a cure for one of the most dangerous illnesses of all time,
>
> on the other hand she's in the medical field and again just like the brain
>
> surgeon she might be able to reach out to many resources for help

Rosy: that's a good point. We would say 'reach many resources/access many

resources' rather than 'reach out'

The example above showed that near the end of the text-based interaction, Majd did not only manage to produce more fluent output, but also much longer utterances compared with the previous sessions.

The saliency of learners' production was another element that triggered pauses during the sessions. That is to say, the learners were seeing their messages on the screen, and hence they spent a few seconds reviewing and editing these messages before sending them to their partners.

E.g., Majd: {So we have Linda who has been (4 seconds pause) away from the habits

of studying and we} (3 seconds pause)   (deleted output)

So I think the problem is that Lind can't really find her way back to studying after being

away from it for 7 years and that she has a lot of responsibilities with being a mom and

all of that   (amended output)

Commenting on the hesitations and revisions identified before responding to his partner, Majd said: "I was double-checking whether what I wrote was accurate and meaningful. I also aimed to exclude all the bits that I did not feel very confident about and produce sentences which resemble the ones my partner was sending, in terms of grammar and structure". This was expected given the affordances of the written online mode (e.g., time), which made it possible for the learners to view their messages on the screen and edit their output as deemed necessary. Lara, on the other hand, had a different justification for the pauses identified by analysing her online performance. She traced back most of the pauses that she made in the first couple of sessions to the interference of studying another language; by the time of the study, Lara had

been studying French for three years and therefore she was struggling sometimes when typing in English. She said:

> "I was thinking that maybe you would say that I am dyslexic or something, especially when it comes to pronouncing/spelling words like decision, literature, etc. I should admit that it took me some time to switch back to English. This might justify the delay in my responses and why I was hesitant when typing my messages; I felt like I needed more time to process the language and form an appropriate answer due to the interruption that I had while learning another language."

Lara mentioned other reasons for pausing during the online sessions: she mentioned that she preferred to deliver meaning in the most appropriate way in terms of grammar, structure, spelling, and word choice, especially because her partner was a native speaker. Thus, one could assume that learners who endeavoured to produce grammatically correct and complex structures, seemed less concerned about the speed or the fluency of their production.

The situation was different for Ameen, although he was also pausing very often during the written sessions. The lack of rapport and chat protocol between him and his partner was causing a lot of pauses, hesitations, and confusion on Ameen's part, who was already a bit apprehensive about interacting online with someone he has not met before using a second language. Once the sessions started, Rami was acting more like a teacher, i.e., being very formal and leading the conversation with Ameen. There was also not much organisation in terms of turn-taking as both interlocutors were texting at the same time and hence Ameen was pausing either because he was waiting for his partner to finish or because he was reading a message that Rami already sent while Ameen was typing another. In addition, as Ameen was typing his messages, his partner was not offering him the chance to finish, rather he was interrupting him by asking yes/no questions.

E.g., Rami: Do you see any books on the shelf?

Ameen: yeah

Rami: Ok cool

Do you see a TV stand with a TV on top of it to the left of the picture?

Ameen: yeah

Rami: Can you see a table with two candles on it?

Ameen: {in the (4 seconds pause) left of the picture there is a table and (5 seconds pause)2 ca} (6 seconds pause)

yeah

In this example, Ameen was, in fact, finally writing something about the table and the candles, yet as his partner had already talked about that, he was obliged to delete this sentence and move to talk about something else. This indeed affected his production and led to more pauses. During the interview, Ameen commented:

"I am not sure I was benefiting much from these sessions as my partner was the one doing most of the talking and I was not getting the time and the space to practice my L2. I also felt that there was some sort of awkwardness while interacting with Rami, haven't got a clue where that feeling came from, but I noticed that there was a lot of unexplained silence and pauses."

Compared with the other learners, Rima, who was in the control group, made fewer pauses across the written sessions. This could be because she was mostly typing short sentences as shown in the example below where Rima was describing the content of her picture to find out the differences between her version and her partner's.

Figure 14: spot the difference

Picture A
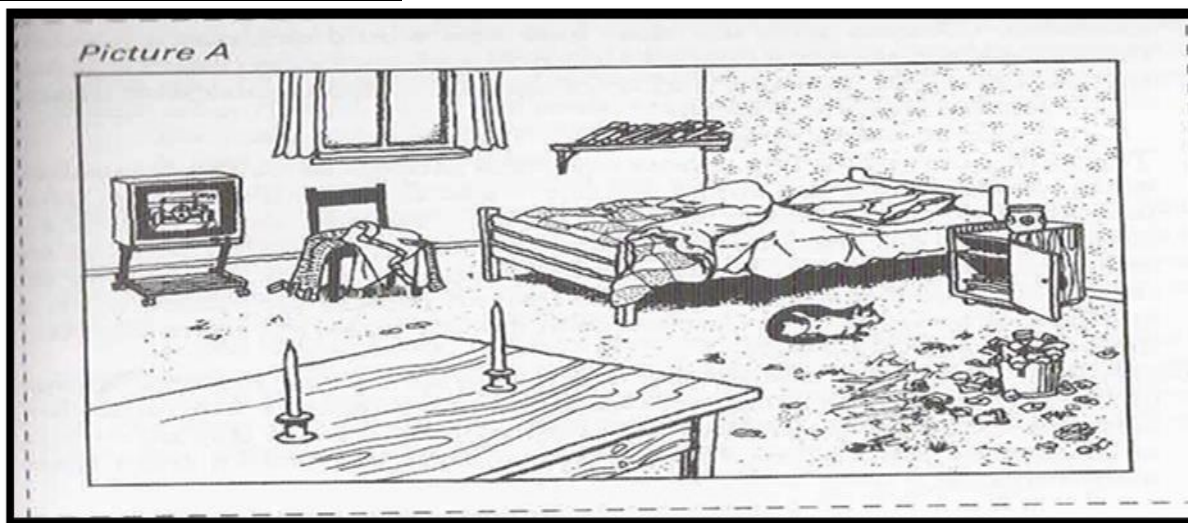
    E.g., Rima: It's a bedroom

      Rana: Ok

      Rima: Untidy bedroom

        There is a bed and a table (pause) on of it 2 candles

      Rana: Here I have two candles on the table also

      Rima: Ok

        There is a bedside with an open (pause) board

      Rana: How about the bed cover?

      Rima: It is untidy

The pauses, as Rima reported, were also because she was only focusing on the content of task as well as her word choice. Rima commented:

> "I was reviewing and revising my ideas and trying to think in English. While writing I was worried about how to spell new words; sometimes I refrained myself from using some words mainly because I am not 100% sure how to spell them. This was another reason for the pauses I made during the sessions".

To sum up, it is highly unlikely that all the pauses the learners made during the written sessions were a clear-cut indication of disfluency or due to their inability to express themselves; rather,

learners might have simply needed time to think of the questions that had been naturally prompted during the session as a result of the interaction between the interlocutors. Pauses generated in the narrative tasks, in particular, were primarily because the learners were re-reading their partners' description of the pictures, taking some time to think of a plausible scenario for the story and figure out the right order of the jumbled pictures. Additionally, learners having different typing skills was another contributing factor that could have impacted how frequently the learners paused per session.

Having analysed the written online performance of the learners by closely examining their L2 production across different CAF measures, the findings show that the treatment received by the planning group (i.e., planning time, guided instructions, and feedback on learners' errors) did have some positive effect on their production of relative clauses. However, no noticeable progress was found in terms of the accuracy and fluency measures of learners' output. Hence, when we attempt to account for gains, we need to consider a number of potentially contributing factors, such as mode of interaction, the task type, the amount/type of feedback provided by NSs/STs, the effects of another language (interference), the rapport between the learners and their partners, as well as learners' typing skills. These seemed to have negatively impacted the extent to which the learners were able to produce an output that is complex, accurate, and fluent at the same time. The next section addresses learners' performance across the audio sessions.

5.2.2 Voice chat

The analysis of learners' chat recordings during the oral mode was carried out in the same format as the written mode with the aim of uncovering any factors that could have influenced their oral performance. Similar measures were employed to operationalise CAF and understand the rationale behind learners' production or otherwise of relative clauses, errors,

and pauses per session. These are addressed and interpreted accordingly in the following paragraphs.

5.4.2.1 Complexity measure

The complexity results revealed that learners' employment of relative clauses differed throughout the voice chat sessions (see **table 49**). Apart from Majd, who produced more language and consequently more relative clauses, neither of the other two learners working in the treatment group managed to generate relative clauses during the first chat session.

Table 49: Complexity in the audio sessions

Tasks	Groups	Pairs	Relative Clauses
Spot the difference	TG	Majd & Rosy	16
		Lara & Kate	0
		Ameen & Rex	0
	CG	Rima & Rana	0
Decision-making	TG	Majd & Rosy	5
		Lara & Kate	6
		Ameen & Rex	2
	CG	Rima & Rana	3
Narrative	TG	Majd & Rosy	6
		Lara & Kate	0
		Ameen & Rex	2
	CG	Rima & Rana	0
Decision-making	TG	Majd & Rosy	5
		Lara & Kate	9
		Ameen & Rex	5
	CG	Rima & Rana	6
Decision-making	TG	Majd & Rosy	3
		Lara & Kate	5
		Ameen & Rex	3
	CG	Rima & Rana	2
Decision making	TG	Majd & Rosy	7
		Lara & Kate	8
		Ameen & Rex	2
	CG	Rima & Rana	4

A close analysis of the chat recordings revealed that learners' production of more relative clauses and complex structures could be attributed to how often NSs/STs incorporated these forms into their own production during the task. This might explain the high number of relative clauses integrated by Majd, particularly in the first audio session where the interlocutors were working on the spot-the difference task.

Figure 15: spot-the-difference 2
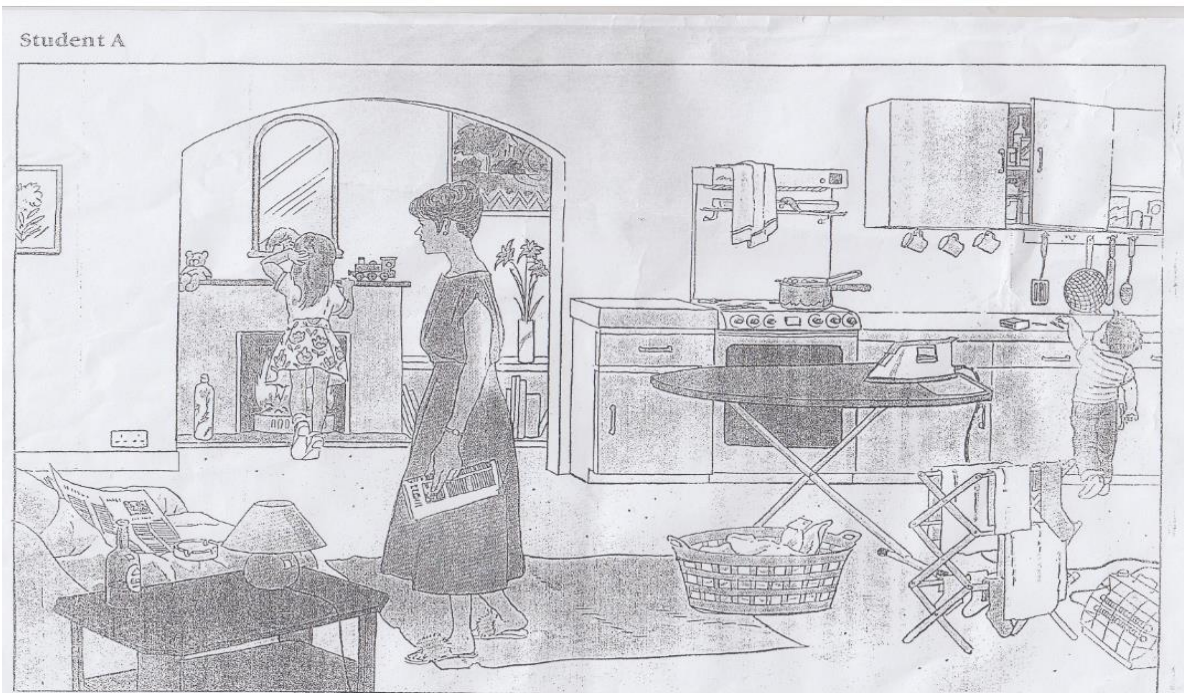
> **E.g.,** Majd: the bottom left corner, there is a couch **which has a newspaper on it**
>
> Rosy: yes, I also see a couch with a newspaper
>
> Majd: next to it to the right, there is a table **that has a lamp on it as well**
>
> Rosy: yes, another difference. I think on my couch I see a cigarette and an ashtray
>
> Majd: there is an ashtray but there is no cigarette
>
> Rosy: ok, so you see an ashtray **which doesn't have a cigarette in it**
>
> Majd: yes, exactly

Majd consistently reported that he was imitating and replicating his partner's style by producing similar structures, with the aim of producing more native-like language. However, as the sessions progressed, he was found to produce noticeably a fewer number of relative clauses (see **table 48**). This could be due to the fact that Rosy was providing less feedback during the oral mode compared with the written equivalent. During the interview, Rosy mentioned that she found it inappropriate to deliberately interrupt the flow of the conversation and trigger the learners to incorporate complex forms into their production:

"Well, I did like being able to hear their voices and hear sort of how they formulate things in a spoken mode, but when you are not in front of them like able to write things down in front of them or even to text while speaking I do find it hard to sort of interrupt them."

Additionally, learners' hesitancy to use these forms, even though they were instructed to do so, was traced back to the pressure they experienced due to the instantaneous nature of oral production, which perhaps impeded their ability to produce complex output. Ameen, for example, reported that the lack of thinking time during the oral sessions made him feel a bit nervous, saying that: "that was tense, there were times where I was unable to retrieve the appropriate vocabulary to describe the picture to my partner, let alone remember to employ complex structures like relative clauses". Besides, the interaction between Ameen and his partner was poor as it was mostly based on short/simple sentences, and hence opportunities for producing elaborated utterances, especially in the first session, were very unlikely.

E.g., Ameen: to the right there is a train

Rami: sorry there is what?

Ameen: there is a train

Rami: ok, a toy train

Ameen: to the left there is a small bear

Rami: teddy bear, ok

Ameen: sorry?

Rami: a teddy bear, we are talking about a teddy bear

Ameen: yes, teddy bear

One may be tempted to assume that the planning time learners had prior to the oral sessions as well as the treatment they received in the written mode did not result in any developmental gains in terms of complexity. However, in the light of the qualitative findings (i.e., learners'

production of more relative clauses, fewer errors/pauses) and the interpretations discussed earlier in the text chat section, this assumption is probably inaccurate. Rather, the learners failed to make good use of the planning time they had been offered before the sessions because they were generally unfamiliar with the note-taking strategy. The fact that most of the NSs/STs were not fully following the given instructions (e.g., giving feedback on learners' errors, pushing learners to employ relative clauses) could have also potentially affected the outcome of the treatment received by the learners in this study. That is to say, the lack of feedback on learners' output could have hindered learners' production of prolonged and complex sentences. The type of task was another reason that inhibited the integration of relative clauses as reported by Lara. She said that she felt more relaxed when working on decision making tasks compared with the other task types and therefore managed to produce more complex structures in the former; indeed, this was evident in the fourth session as shown in the table above. One could also presume that there was a link between how relaxed and how motivated the learner was and the number of relative clauses they produced per task. For her part, Rima in the control group succeeded in producing some relative clauses (using **that** mainly) while she was interacting with her partner. Again, these were primarily perceptible during the decision-making tasks which enabled the integration of a greater variety of language forms. Discussing the pros and cons of moving to another country and deciding on whether or not to take that opportunity, Rima wrote:

> **E.g.,** Rima: I think it's up to the the guy **that's you are talking about**. Maybe not all of them they love to travel, okay other than the countries of Saudi Arabia and things like that. You know some people get better salaries and they get better life, better lifestyle I mean. That's it. so just people **who love to live in these conditions could like it**.
>
> Rana: ok

Rima: Just to get more money, I think.

Rana: yeah

Although Rima used relative clauses twice in the excerpt above, these were not always error-free. Having discussed the complexity findings of the audio sessions, the following section turns to accuracy.

5.2.2.2 <u>Accuracy measure</u>[23]

Learners' errors during the voice sessions were counted and logged for further analysis, whether these were related to grammar, content, pronunciation, and/or word choice. Akin to the text chat, inconsistencies were uncovered with reference to the progress made by learners working in the treatment group throughout the sessions (see **table 50**). There was no reduction in the number of errors the learners were making while interacting orally with more proficient L2 speakers. Rather, more errors were identified as the online sessions proceeded compared with the first session. This could be explained by the different type/content of tasks the learners worked on; presumably, a decision-making task would generate more elaborated language than a spot-the-difference task, and apparently more errors as well. Thus, making more errors could also mean that the learners were speaking and interacting more during the session, which in one way or another may suggest an improvement in their L2.

---

[23] I have **bolded** learners' errors in this section.

Table 50: Accuracy in the audio sessions

Tasks	Groups	Pairs	Number of errors
Spot the difference	TG	Majd & Rosy	46
		Lara & Kate	34
		Ameen & Rex	33
	CG	Rima & Rana	48
Decision-making	TG	Majd & Rosy	27
		Lara & Kate	68
		Ameen & Rex	44
	CG	Rima & Rana	86
Narrative	TG	Majd & Rosy	20
		Lara & Kate	65
		Ameen & Rex	26
	CG	Rima & Rana	37
Decision-making	TG	Majd & Rosy	49
		Lara & Kate	49
		Ameen & Rex	28
	CG	Rima & Rana	111
Decision-making	TG	Majd & Rosy	40
		Lara & Kate	62
		Ameen & Rex	25
	CG	Rima & Rana	98
Decision making	TG	Majd & Rosy	51
		Lara & Kate	62
		Ameen & Rex	40
	CG	Rima & Rana	60

Likewise, producing fewer errors could not necessarily be equated with gains in learners' accuracy; it could be simply because the learner was not speaking much, and/or was mainly producing concise sentences as they were not feeling sufficiently relaxed/comfortable to elaborate more on their ideas. When asked why he was not giving more details during the sessions, Ameen said that:

"It was very challenging to speak for a long duration, without having time to think or edit my language. Therefore, I was mostly either repeating myself or producing short sentences to avoid making more errors."

E.g., Rami: How would you stand for this person?

Ameen: well, I will **<u>defined him</u>**. Be his friend, like that

Rami: Okay, so basically you will stand next to this person who is getting bullied and you won't hang out with your friends anymore?

Ameen: yeah, maybe

Rami: Okay so how do you think your personal values and believes affect your decision?

Ameen: (long pause) and no answer

Rami: Let's say as a culture we have something called you know dignity and you know whenever we see someone is getting bullied it's a Syrian thing in our culture when you see someone getting bullied you can't just stand aside and just you know. Seeing him getting bullied and you can't do nothing. You have to do something about it. You have to help this person, is that correct?

Ameen: yeah, that's true 100 percent

Rami: yeah anything else to add?

Ameen: well, when you imagine that when you put yourself in his position and imagine what **<u>the others do</u>** to you, I think **<u>we would</u>** do something to that person.

The immediacy of the oral mode and the fact that no feedback was provided on learners' errors could have also affected the quality of the language they produced per session. Lara, for instance, was very motivated and excited to speak, yet her sentences were not error-free; she was persistently making grammatical errors which, though they were not impacting the delivery of her messages, indicated that there was an existing problem in her L2 uptake. Her partner, nevertheless, had chosen not to draw Lara's attention to these errors as long as she

could understand the gist of Lara's messages. We see evidence of all of this in Lara and Kate's discussion of the following task:

> *You've just started working for a software start-up. One day you have a great idea for a new app, which you believe is going to revolutionize the industry. You share it with James, your co-worker. The following day, you find out James had presented the idea to the boss, claiming it was his. Needless to say, the boss loved it, gave James the credit and the promotion, and the app is going to be produced soon.*
>
> ***Together with your partner <u>discuss the reasons that you think might make James do what he did</u>, <u>think of ways/solutions to convince the boss that the idea was yours.</u> <u>What would you do if all your attempts to convince the boss were in vain?</u> Are you to blame for what happened? You might also discuss any learned lessons?***

The interlocutors wrote:

> **E.g.,** Kate: you can improve it and start producing it yourself or give it to another
>
>   company
>
>   Lara: yeah, I would do so but like, **I'll have the courage**, if this **happens** to me,
>
>   I would like be shocked and never trust people ever again and quit
>
>   everything. But thinking about it like from a distance like what, **if the**
>
>   **damage is done what he can do to fix it**? Nothing. So **leave it alone** and
>
>   try to make things better **in another place**.
>
> Kate: either that or if the boss doesn't listen to you, there must be somebody else
>
>   that the boss does listen to so you could go to them, like if there are other
>
>   managers or you know sometimes even the boss of the company isn't the
>
>   highest person so they might not be like, they might have somebody else that
>
>   they answer to

This example showed how the NS, Kate, chose to ignore Lara's errors and proceeded with the conversation. Hence, it seems that measuring the accuracy of learners' production throughout the voice chat sessions is not a straightforward process as there were different factors that need to be considered, including the mode itself, the task type, and the feedback given to the learners. The following section discusses the fluency of learners' production during the text-based chat.

5.2.2.3 Fluency measure

In line with the analysis of the written chat, learners' pauses (filled/unfilled) throughout the oral sessions were counted and logged for further analysis, aiming to provide interpretations for learners' pauses, i.e., what made the learners pause during the sessions? Unfortunately, the length of pauses could not always be accurately calculated because of the intermittent internet connection and the fact that the speakers' turns were sometimes delayed because of poor connectivity. Long and more frequent pauses were sometimes noticed when the learners were trying to answer a question that arose spontaneously during the conversation with their partners, i.e., a question they did not think of while planning for the task. This is demonstrated in the following example, when Rosy chose to expand the discussion with Majd and started to ask further questions on the pros and cons of moving to another country.

E.g., Rosy: ok, and how about for keeping in touch with the people in your home country?

Majd: well, it would still be easier to keep like (pause) in touch with your (pause) people back in your home country (pause) through these apps coz like it's (pause) less stressful than going to (pause) the mail office writing letter and like (pause) going back and forth to the mail office. So yeah, I think through these apps it's easier whether to keep up with your friends and family back in your home country or meet new people in your (pause) in the country that you're moving to.

Commenting on the pauses identified in the example above, Majd said:

"There were a few instances where I found myself pausing a lot to kind of find an appropriate answer to the questions that Rosy was asking during the sessions and which I did not think of earlier. This example was absolutely one of them."

Table 51: Fluency in the audio sessions

Tasks	Groups	Pairs	Pauses
Spot the difference	TG	Majd & Rosy	97
		Lara & Kate	55
		Ameen & Rex	19
	CG	Rima & Rana	54
Decision-making	TG	Majd & Rosy	64
		Lara & Kate	134
		Ameen & Rex	28
	CG	Rima & Rana	110
Narrative	TG	Majd & Rosy	54
		Lara & Kate	93
		Ameen & Rex	33
	CG	Rima & Rana	89
Decision-making	TG	Majd & Rosy	114
		Lara & Kate	148
		Ameen & Rex	50
	CG	Rima & Rana	149
Decision-making	TG	Majd & Rosy	62
		Lara & Kate	117
		Ameen & Rex	53
	CG	Rima & Rana	146
Decision making	TG	Majd & Rosy	110
		Lara & Kate	133
		Ameen & Rex	46
	CG	Rima & Rana	123

As shown in **table 51**, the learners paused more as the tasks increased in their complexity. Initially, the qualitative analysis of the chat recordings revealed that cognitively demanding tasks, especially those that were based on free conversation, required more thinking time and therefore more pauses were apparent. During the interviews, the learners were asked to recall their thoughts about all the identified pauses and hesitations in their online performance; different explanations were provided. For example, learners reported that most of the pauses produced while working on picture-based tasks were either due to them going back to the pictures to review their content, and/or sometimes because they were thinking of their word choice: i.e., they were unsure about the right vocabulary to use in a specific context.

Figure 16: spot the difference 2 (a & b)

  E.g., Rosy**:** Ok, I only see one item hanging. Can you tell me what items are hanging from

   the hook?

  Majd: so there is this (pause) circular (pause) sort of like dotted (pause) ban like

  Rosy: I'm not sure, like a pan?

  Majd: It's not a pan. It's like (pause) the thing that you use to like (pause) get out the

   fries?

  Rosy: oh is it like spoon with holes in it or like a colander that is sort of, it's like a

   bowl with holes in it and you drain things with it?

  Majd: yes, it's like (pause) a bowl which has holes in it

  Rosy: ok that's like a colander. That's what we call it

  Majd: colander ok, noted

E.g., Majd: Ok, (pause) up to the left (pause) Next to the girl to the right, upper right

there's some sort of a painting of trees as well

Rosy: yes, I also see a painting that has a tree and it is a sort of landscape painting

Thus in the example above, the vocabulary item that Majd wanted was 'colander', but because he was not familiar with it, he was forced to paraphrase so that his partner could understand what he was talking about and provide him with the right word. Commenting on the pauses in the second example, Majd said: "I was looking and relooking at the picture to determine what preposition of place to use. I just wanted to be precise in my description, so that I do not cause any misunderstanding for my partner."

Pauses and hesitations were also caused by learners focusing on their sentence structure as well as trying to incorporate new expressions that they heard from their partners during the decision-making tasks. Discussing the reasons that might have made James (a co-worker in a software start-up company) steal the idea for a new app, the interlocutors wrote:

E.g., Rosy: so he stole the opportunity and do you think he is the kind of person who

would do anything **to get ahead in life**?

Majd: maybe he is that type of person

Rosy: okay and why do you think he wasn't like potentially fussed about betraying

his colleague? Do you think he did it on purpose?

Majd: well, like I said maybe he (pause) needed the promotion money or something

so for, like personal reasons so he (pause) took advantage and he stole the idea.

Maybe, maybe he just like (pause) try (pause) is trying to use like (pause) me

for like for my ideas or something too. Like you said (pause) **to get ahead** or

to get attention from the boss.

In his interview, Majd said: "I have the idea in mind but still need to think of an appropriate way to deliver the meaning to my partner in terms of grammar and structure." Besides, when asked about the pause that he made when integrating the expression that was previously used by his partner ('to get ahead'), he commented:

> "This is maybe due to the lack of practice; we rarely have the opportunity to speak to someone in English, and consequently we don't get the chance to integrate these words into our production very often. To be honest, I knew some of the expressions that Rosy was using during the sessions. However, the fact that we were working on different tasks, and that there were maybe too many aspects to focus on, made me bit hesitant when reusing these expressions."

To put it another way, as the tasks' content varied throughout the sessions, it was very challenging for the learners to appropriately reuse these expressions.

Most of the pauses generated in the narrative task, however, were because the learners were thinking about how to unjumble the pictures and trying to make sense of the description provided by their partners to achieve the task's goal. However, there was a case where one of the learners, Lara, seemed confused, thinking that she should not look at the pictures while describing their content to her partner. This then resulted in a lot of pauses in the learner's utterances as she was trying to remember what she saw earlier in the pictures:

> **E.g.,** Lara: OK, (pause) the first one, (pause) there were a lady and a man were walking down the street. (pause) I think they saw a pet shop (pause) and (pause) the woman admired (pause) a black cat (pause) and this cat was (pause) for sale.
>
> Kate: okay, what are their expressions? Are they happy or?
>
> Lara: I can't tell. I think they were happy. They were like smiling but the lady was (pause) the most happy
>
> Kate: ah, ok

Commenting on the pauses she made in the example above, Lara said: "not sure why I thought that I can't keep the pictures while talking with my partner and so I paused because I was trying to remember what I had in each picture".

It is worth mentioning that throughout the voice chat sessions in particular, Lara was very motivated to learn, yet she was fully aware of the interference of studying/learning other languages and how this had negatively impacted the speed of her English production; hence, this might possibly also partly explain the pauses that she was making. Reflecting on her online performance, Lara commented:

> "I also faced some challenges when speaking, influenced by my French studies, and the fact that I was also trying to learn Turkish (enrolled in a Turkish course). You could imagine how the three languages were mixed up in my mind as each language has different rules in terms of grammar and sentence structure. This might justify why I was pausing more often when speaking to my partner; I felt like I needed more time to process the language and form an appropriate answer."

The situation was, however, different for Ameen who did not benefit much from the planning time; he was not taking full advantage of the ten minutes given prior to the online performance and ended up writing only a couple of notes. This probably explains why he was pausing very often during the audio sessions. Ameen was reluctant to speak much throughout the online sessions in order not to make more mistakes or sometimes because he did not have enough words to express his ideas. He mentioned that he paused a lot while talking with his partner because he had to think of what to say and how to say it: "after all, I was talking with someone who speaks English perfectly, and hence I did not want to make more errors". He therefore referred to the lack of thinking time during the voice chat, compared with the text chat, as one of the shortcomings of this mode. Additionally, the chat excerpt below demonstrated how the

lack of rapport between Ameen and his partner also caused a lot of pauses and hesitations on Ameen's part.

E.g., Rami: …So the next question is why do you think some people like to bully others?

Ameen: Maybe to show them that they are stronger than them or (pause) show who has the power to do (pause) anything they want. (long pause) I think this is maybe things to do (pause) to make someone bullying the others

Rami: Yes true or maybe they are facing some issues at home, Or their parents are bullying them so They think they should bully someone else

Ameen: Yeah that's true.

Rami: So what do you think the long-term and short-term effects of bullying?

Ameen: For short term it would be (pause) depression, hate themselves (pause) for the person who's bullied (pause), lack of self-confidence (long pause). may be the long-term (pause) consequences maybe commit suicide

Rami: yeah, correct

So would you try to convince your friends to stop that action and how?

Thus, as opposed to the main objective of the task that is based on the interlocutors discussing a certain topic/making a final decision in order to give the learners the chance to produce more elaborated output and get them exposed to the more advanced input produced by NSs/STs, Rami was just asking the questions that came in the task description. Rami was not offering any feedback on Ameen's errors either, which was why perhaps Ameen reported that he felt that he was doing the task alone. In the second session, for example, there were many instances of total silence. Rami, however, traced what was happening during the sessions back to Ameen's character:

"dealing with a shy person online without seeing him F2F is really challenging. I was trying to give him some space, pushing him to speak, but unfortunately, he was not very

responsive. I also chose not to correct him very often as I did not want to put more

pressure on him".

On the other hand, Rima, who received no planning time and/or guided instructions regarding her language use, was pausing more often than the planning group in the last three sessions. This happened as she was talking more and producing longer turns compared with the previous sessions. Rima reported that she got used to the online mode and the different task types, and that she felt more confident over the sessions. When asked what she was thinking of each time she paused, Rima replied:

"I was primarily concerned about my word choice; due to the instant nature of oral interaction, I was a bit struggling with recalling the words that suit the given context. This absolutely slowed down the speed of our interaction, yet I think it was better than saying nothing, which I definitely did not want it to happen during the session".

Hence, unlike the learners who worked in the planning group, focusing on vocabulary rather than her sentence structure appeared to be the primary reason behind the pauses Rima made in the audio chat sessions.

To recap, a lot of pauses were identified in learners' production during the Skype calls. These, however, did not always indicate lack of fluency; rather they were sometimes an indication that the learners were talking more compared with the previous sessions. Additionally, the instantaneous nature of oral interaction meant that the learners might have needed time to think of the questions that had been naturally prompted during the session. Having analysed the oral performance of the four learners by closely examining their L2 production across different CAF measures, the findings reveal that the treatment received by those who worked under the guided-planning condition (i.e., being provided planning time, guided instructions, and feedback on their errors) did have some influence on their L2 production. This involved what aspects of their language they chose to focus on (grammar, content, and/or word choice) which

in turn affected the number of errors and pauses generated by each learner per session. Thus, the online interaction was sometimes efficacious in promoting L2 development. However, learners' responses to the given prompts during the interviews uncovered several other factors which perhaps negatively impacted the efficacy of online interaction and caused the inconsistent results in the current study, including learners' motivation/anxiety level, the rapport between the learners and their partners and the type and amount of feedback/correction their more proficient partners were prepared to give.

Moving on to the final research question, the following sections discuss participants' responses during the interviews and elaborate more on their attitudes/perceptions towards the online experience.

5.3 RQ5: What are the attitudes and perceptions of the participants towards the online experience?

Stimulated recall interviews with the learners following each chatting session and semi-structured interviews with all the NSs/STs after they had performed the online tasks served as the main data source to answer this research question. The results revealed that almost all the participants appreciated this experience, enjoyed the online sessions as well as the different task types, and saw chatting as both fun and beneficial for language learning. They also talked about the different features and affordances of SCMC that might help them improve their English. Some negative comments were reported as well. These responses are described below accordingly.

5.3.1 Learners' Attitudes

Starting with the Syrian learners, this section will be divided into learners' attitudes towards the online experience including planning time, task type, mode of interaction, their attitudes towards their interlocutors as well as the online experience as a whole.

5.3.1.1 Planning time

Unfortunately, not all the learners in the treatment group made good use of the planning time given prior to each chatting session, especially in the first couple of sessions where the learners were wasting their time typing full sentences rather than short notes that represent a specific idea. Perhaps this happened because most of the learners were unfamiliar with note-taking skills; even when they were given step-by-step instructions as to how to plan their online performance in the very beginning of the data collection stage, not all the learners were found to follow these instructions, despite the instructions being delivered to the learners in their L1 to try to minimize misunderstandings. Hence, learners differed in terms of how they perceived the benefits of the planning time. Some learners viewed the planning time as unhelpful since, as mentioned earlier, they preferred to organise the ideas in their head rather than commit them to paper. On the contrary, other learners found the ten minutes' planning time prior to the session very beneficial; when asked what she thought of the planning time, Lara responded:

> "I liked the idea of notetaking very much; it was a very practical and stress reliever strategy, especially during the first session where I had to talk to a NS, someone who was totally a stranger for me, and I was very anxious about that. I was also anxious about the fact that it's been a while since I last write/speak in English, and that the whole experience was new to me."

There were also learners who believed that the effectiveness of the note-taking stage varied based on the type of the task; they deemed the planning time as necessary when working on cognitively demanding tasks (i.e., narrative and decision-making tasks). Ameen, for instance,

256

reported that he did not see the logic behind the planning time when preparing for spot-the-difference and information-gap tasks; therefore, he was reluctant to take notes prior to the online sessions. He claimed that describing what you can see in a picture that is already in front of you did not require a lot of thinking time to structure your answer, given that you are already familiar with most of the vocabulary items needed to achieve the task's goal. However, he seemed to have a different view as he started working on more complex tasks that necessitated the use of a greater variety of vocabulary items and language structures. Following one of the text chatting sessions, where he had to discuss the circumstances of six patients who were in need of a heart transplant and decide to whom the heart should be given, Ameen commented:

> "Having some time to organise my ideas was very beneficial with this type of task, which was indeed more difficult and required producing more varied output than just saying: *to the left, there is X*, or *next to X, I can see Z*. I reckon that the planning time also facilitated greater attention to the language that I was producing later on while interacting with my partner, given that all the ideas had already been thought of."

Hence, it seems that strategic guided planning offered before the actual task performance freed learners' attentional resources to attend to other aspects of their language during the online performance, particularly when working on cognitively demanding tasks.

5.3.1.2 <u>Task type</u>

Overall, learners' attitudes towards the various task types employed in the current study were predominantly positive. Most of them mentioned that they were looking forward to the upcoming sessions and were very excited to know what kind of task they would be working on with their partners. Majd, for example, commented: "I did like the fact that we had a different task each week, either in terms of their type or content. I believe that sticking to one task type

would have made the sessions very typical and tedious". Lara also found it very helpful that the tasks were sequenced based on their complexity:

> "I noticed that each week, the tasks were getting more difficult, and in fact, sequencing the tasks as such was really wise and sensible. I believe that it would be very distressing to start by working on complex tasks and discuss real problems with someone I've never met before. I think I needed that time to open up and be ready to share my thoughts regarding particular topics with my partner".

When asked to order the tasks based on their complexity, the learners reported that spot-the-difference/information-gap tasks were the easiest in terms of their content and cognitive demands; the learners generally viewed these as ice-breaker tasks which gradually scaffolded them to work on more complex tasks. The narrative tasks, on the other hand, were said to be the most difficult for two reasons: firstly, the learners did not have all the pictures during the sessions as these were equally divided between the partners; and secondly, the learners had to figure out the right sequence of pictures and come up with a comprehensible story based on their partner's description of the other pictures. Lara commented: "this task was really difficult; I was not able to distinguish the characters. Like I could not tell whether the man and the women in my pictures were the same as in my partner's pictures". The challenges faced by the learners as they were working on the decision-making tasks, however, were basically related to the fact that unlike the picture-based tasks, these were based on free conversation. Hence, this meant that the learners needed to produce a greater variety of lexical items, tenses, and sentence structures. They had also to come up with enough ideas to discuss the existing problem with their partners and eventually make the right decision. That was not always easy and straightforward; the topics were mostly taken from real-life situations and thus some raised difficult issues and required more thinking time than others (e.g., the bullying and heart-transplant tasks vs. the smoking task). Commenting on her online performance when working

on the smoking task, Lara said: "this is a very controversial topic and since I am not a smoker, it took me some time to try and think of how smokers feel and what would make them smoke in the first place, so that I could provide a more persuasive argument."

However, Ameen had a different view about sequencing the tasks from simple-to-complex; he preferred to work on tasks which tackled a situation that he could easily relate to his own life and his own specialisation as an IT student (e.g., moving abroad). He noted that he found these tasks simpler than the others as he could easily come up with ideas on the topic and prepare for the discussion with his partner. However, he was less comfortable working on the picture-based tasks and hence he considered these to be more difficult than the other task type. Rima, who also noticed that the tasks were increasing in their complexity, reported that she enjoyed working on each task type.

5.3.1.3 <u>Mode of interaction (written vs. oral)</u>

During the recall interviews, the learners were asked about their feelings towards online chatting via text and voice chat; their responses were mostly positive for both forms of chat. They talked about the distinctive features of each mode and how they thought these had helped them improve their L2 performance. Majd, for example, mentioned that the time span he had during the written chat to revise and edit his sentences (though relatively short), made him more attentive to his grammatical errors, word choice, and sentence structure. He also referred back to the feedback that he was receiving from his partner and the fact that he managed to take separate notes about his partner's corrections as he was seeing these on the screen. This is indeed a distinctive feature of text chat. Lara, on the other hand, believed that text-based chatting is less threatening compared with other modes of interaction (e.g., F2F or voice chat), especially "when it comes to talking for the first time to a complete stranger using another language". Rima also thought that text chat made her feel relaxed due to the lack of instant

verbal interaction:

> "After all, I was just setting behind the screen, and hence I was more confident and daring to express myself. The slow pace of the written mode also helped me to focus on the task that we are working on".

It is worth mentioning, however, that during text chat misunderstanding could frequently happen as the interlocutors were not seeing and talking to each other orally; that is to say, exchanging feelings in text chat could be misinterpreted because of the lack of face to face context. This was in fact reported by Majd following the first texting session; he mentioned that he was feeling "intimidated" when his partner was commenting on his language use. These feelings disappeared though as soon as the oral sessions began. Majd commented:

> "Having heard my partner's voice, I realised the calming and friendly tone that she used when she was correcting my errors. Perhaps this was one of the flaws of text chatting; you see words that lack emotions, and therefore, feelings in text-chatting might get misinterpreted".

Another problem of text-chatting as described by Lara was the overlap in turn-taking; this frequently happened when she was writing something and before sending it, she noticed that her partner was asking a question about something different. In this case, Lara chose to delete her message and respond to her partner in order to maintain the flow of the conversation.

Reflecting on their performance in both modes, the learners showed a preference for voice chat over the text-based equivalent. This was predictable since most of the learners were aiming to practise and improve their speaking expertise. The learners believed that text chat would assist language users to organise their ideas and attend to form, and hence would be a good step to start using and learning that language. They also believed that starting with written chat was

beneficial; it allowed them to become acquainted with their partners and helped them build a good rapport with each other before they moved to voice chat. Majd commented that after written chat: "I reckon that now she knows me better in terms of my proficiency level, my weaknesses, and so on". Lara also talked about how the written sessions gradually prepared her for the oral sessions:

> "The written sessions were like an ice breaker; they enabled me to get to know my partner a little bit more and avoid the awkwardness of immediately speaking to someone I've never met or spoken to before, especially because we were interacting online rather than f2f."

The learners were generally excited and looking forward to starting the voice chat sessions. Still, they were a bit anxious about their ability to express themselves to their partners and the fact that, unlike for the written chat, they did not have time to think of what to say and attend to different aspects of their performance. Majd said:

> "Can't wait to start the oral sessions as I have never had the chance to speak with an English native speaker before. Nevertheless, just thinking that I will be talking to someone in English for like twenty minutes, made me feel a bit nervous."

To put it differently, the learners knew that the oral interaction would be more instantaneous and accordingly more challenging than the written chat. Thus, they were worried about their ability to produce meaningful and grammatically correct sentences. Similarly, Rima was also worried that she might be too pressured to understand what her partner was saying and/or be incapable of finding the right words to interact properly during the session. The learners, however, mentioned that they started to feel more relaxed and confident as the oral sessions proceeded and that they were sometimes asking questions, negotiating the problem and leading the discussion with their partners.

5.3.1.4 <u>Attitudes towards their interlocutors</u>

Having a more proficient speaker to chat and practise their English with (i.e., a native speaker of English or someone whose English level is significantly high) was the primary reason behind learners' choice to take part in the current study. Most of the learners had positive attitudes towards their partners, whether the latter were NSs or STs. When asked about her feelings toward interacting with her partner, Lara responded: "I was lucky to get the chance to talk to a NS. My partner was very friendly and supportive; even though we have not met before, I think we got on well together". She also added:

> "When I first knew that I'd be talking to a NS, someone who is British, I got a bit worried; I heard that British people are very formal and use lots of idioms while speaking. Besides, it's been a long time since I last practised my English with anyone, let alone with a NS. Therefore, I did not feel confident enough to write or speak in English at the beginning."

Rima, who interacted with a ST, had a similar opinion about her partner: "I learned a lot from Rana; she was very friendly and helped me express my thoughts and put them into words. This, in fact, meant a lot to me". Hence, this suggests that the learners were noticing and appreciating the feedback provided by their partners. Lara reported: "I admired the way she corrected my language flaws; she was very considerate as she corrected me indirectly, like by rephrasing what I said, so that I don't feel intimidated or something". Majd, however, seemed to have a different opinion about his partner's feedback, which was mostly explicit. Following the first chatting session, Majd commented:

> "I was feeling intimidated and I felt that the conversation was very formal; whenever I made an error, she was immediately correcting me by saying; 'in English, we say it like that…'. This made me very nervous indeed, yet I was trying to attend to what she was saying and what I was writing as well".

Hence, perhaps learners' perceptions of their partners' L1 status made them more attentive to their language use and the given feedback. This feeling of intimidation, however, changed gradually, especially as the voice chat sessions began; Majd became more relaxed about receiving feedback from his partner. He mentioned that after he spoke to his partner, he received her feedback differently as she was using a very friendly tone. He also admitted that he benefited a lot from the comments she made on his errors and noticed that he was really progressing. The learners were also pleased as they kept the same partner throughout the online interactions; someone they had already built a good rapport with. This was reflected in Rima's comment: "knowing Rana prior to the voice sessions helped me very much and I felt very relaxed and comfortable as soon as we started our first talk." It is worth mentioning though that some other interlocutors (e.g., Ameen) had less positive experience as they did not seem to get on nearly as well with their L2 partners. During the interview, Ameen commented:

"There were times where I felt like I was doing it all by myself, yet there were other times where Maher was taking the lead and not giving me the chance to say anything. Like for example, when we were working on the narrative task and we had to describe the pictures that we each have to each other to be able to create a story at the end, I did not feel that the information he was giving was very helpful. Like he was only giving me a general and very brief description, and then when I started to kind of write up my own version of the story to share it and discuss it with him, I was shocked to see that he already sent me his own narration and just ended the session afterwards. That was very irritating, to be honest."

5.3.1.5 <u>Attitudes towards the online experience</u>

When asked to evaluate and elaborate on their attitudes towards the online experience, SLs' responses fell into one or more of the following three areas:

I. Entertaining & Unique

The learners mostly enjoyed the online interaction as well as the tasks they were working on. This was noted in Rima's comment:

> I was enjoying every single minute of the online session; the topics that we were discussing were very entertaining. I was not even worried about how many ideas I have got to discuss with my partner, given the time we had. On the contrary, I felt that the sessions were going very fast, and I just hoped they lasted for a longer time."

Rima also mentioned that the online experience was unique as she had never previously had the chance to practise her English with more proficient L2 users. She noted: "Unfortunately, it's not something that you can always get, you know, especially in our context."

II. Helpful & Exciting

SLs commented on how helpful they found this experience, either in terms of the written nature of SCMC that drew their attention to errors, and/or the affordances of the oral mode that enabled them to practise their speaking skills with more proficient L2 users. Majd commented:

> "though I was very excited to take part in this study, I tried not to put high expectations regarding the benefits that I would be gaining out of it. I just did not want to get disappointed if no change happened. Yet, I should admit that the benefits I gained throughout the sessions were far above my expectations."

Lara also noted: "overall, I believe that these sessions were helpful as I noticed a significant progress in my language, and I am really delighted about this improvement. I was not bothered about the errors I was making because I knew that at the end of the day, I was learning something new." Rima found the sessions very exciting since there was a new task to work on each week. As she reported: "this made me forget about any worrying feelings that I might have had prior to online performance".

SLs also emphasised the gains they achieved throughout the sessions due to the online interaction with more proficient L2 users. Majd remarked:

> "I believe that the whole experience was educational in the sense that I learned lots of new things in terms of grammar, new expressions, and so on; all of which I was attempting to employ while interacting with my partner".

He also traced the effectiveness of this online experience back to the feedback he received from his partner throughout the sessions:

> "I think what made these sessions very special was the comments that I was getting from my partner regarding my language use and/or my word choice. At first, I did not have the courage to elaborate on my thoughts as I did not want to make more errors, but then I began to feel more self-confident about using the language. I was thinking that even if I took the risk and employed new expressions/structures, my partner would be there to correct me, and hence the gains would be doubled. Now, I could confidently use these forms knowing that they are grammatically correct and meaningful."

Lara, on the other hand, made a favorable comparison between this online experience and the English courses she previously enrolled in. She commented:

> "Personally, I believe that this is the most successful way to learn a language. Now I am thinking that all the traditional F2F classes/courses that I took, and which were mostly grammar-based, were not as effective as this one. I got the opportunity to practise my English and discuss different important topics with someone whose first language is English. I admit that learning grammar would be necessary if you intended to pursue your academic studies. However, if you want to fluently speak a language, I believe that everyday practice will be much more beneficial. Of course, in the latter case grammatical rules related to this language would be attained subconsciously."

Having addressed learners' attitudes, the second part of this section presents NSs' and STs' perceptions of the online experience.

5.3.2 NSs/STs' perceptions

All the NSs/STs were interviewed at the end of the online sessions to get more insights into their perceptions towards the online experience. Overall, NSs/STs described this experience as interesting, new, and beneficial for language learning. Rosy commented: "Yeah, I found it very interesting. Overall, I felt the experience was positive and I really enjoyed taking part". Melanie was also positive:

> "I really enjoyed it, I thought that it was really fun. I found it quite rewarding and it was really nice getting to talk to people that I would probably not normally talk to. It was also nice to kinda think about how I used my language differently and thinking about how I describe things in a way that's clear and easy for someone else to understand; like someone whose first language isn't English. So that was quite useful."

The participants were mostly unfamiliar with SCMC as a tool for language teaching and learning. However, they thought that participating in the current study was advantageous in the sense that they got to know more about the different modes of interaction which could be employed for educational purposes. Rami stated: "I gained more experience as a teacher since this was the first time I practised online teaching. In fact, most of my teaching experience was based on F2F grammar-based courses".

In addition, almost all the NSs/STs mentioned that they would be willing and happy to take part in future online teaching studies. Anna reported: "if I got the chance to do this again I would definitely do it. I think it's potentially really purposeful and a useful way to learn". Though all the participants (mainly NSs and STs) still believed that F2F teaching is more efficient when it comes to teaching a new language, they viewed the online tool as a

supplementary approach that helped the learners to practise the language and apply what they learn in the classroom in more open contexts. Reflecting on her experience of learning Arabic, Kate commented:

> "I found F2F learning more effective than the online learning. I feel like online learning is a bit more supplementary. If I have somebody in real life, like properly teach me, I can then go and do things online; like if I can build a foundation then I can go into things online".

The following sections elaborate more on NSs/STs' attitudes towards the different modes of interaction, as well as the type of the tasks implemented in the current study.

5.3.2.1 Perceptions towards the modes of interaction

NSs/STs tended to prefer oral interaction over the written (text-based) equivalent as they believed that the former was more natural and authentic, and that they had more language to work with as the interaction was more instant and learners were producing more turns. Commenting on both modes, Esraa said:

> "During text chat I felt like sometimes the meaning was lost and that it was like you are talking to yourself really, whereas when you are on the phone you can hear the tone and you can tell the meaning quicker, I guess."

Hence, as both interlocutors were taking time to finish typing their messages, some NSs/STs thought that interacting via text-chat was a waste of time. However, other participants seemed to perceive this feature (i.e., the additional time available for each interlocutor during text chatting) differently, particularly when addressing learners' errors. For example, Anna said:

> "I probably find things like error-correction easier to do over text-chat because it's slower. So I think that actually having the visual element of the text chat and the fact

that it was slightly slower allows me with such a kind of speedy learner to slow things down, and try to put more language input".

She also explained why she was finding the feedback provided during the oral mode not as effective:

"my partner was such a very enthusiastic and keen talker that he would be talking about something really quickly and in a really engaging manner, and then I noticed oh hang on there was an error that was kind of two sentences ago, but firstly I didn't want to always interrupt his flow, and secondly that it was two sentences ago that he probably forgot he said it."

Similarly, Rosy found it very tricky to do the tasks when she could not type (i.e., when the task is just auditory); she wished to give feedback and correct her partner, but she found this difficult to do in the voice chat mode. She reported: "so the first part where we could type, I felt I was able to keep track a bit better because I could see visually what the learners were writing". She also mentioned that she struggled a bit with her ability to concentrate on getting the tasks done while at the same time trying to address the language issues. This made her sometimes feel a bit confused about her role as a participant and her role as a teacher. Therefore, she suggested that maybe doing things differently "like letting two students work on the tasks together" would enable her to focus more on the language and practise her role as a teacher.

Rosy added that she was looking forward to hearing the learners' voices and how they would formulate their utterances in the spoken mode. However, as she was not seeing the learners F2F and was not able to text them while speaking, she found it hard to interrupt them to correct their errors. Thus, she was more apprehensive about interrupting the learners during the sessions. One of the examples that she presented spoke of what she saw as the unintended consequences of interrupting one of the learners to correct him:

"I did not want to wait too long but I sort of broke his concentration; he lost his train of thought for a bit, so it's always hard to find the balance between disrupting them and always drawing their attention to errors".

This feeling of uncertainty was the same for almost all the NS/ST interlocutors; they found it very challenging to address learners' errors. For instance, Sally thought that correcting learners during the chatting mode was not straightforward due to the overlap in turns. She commented:

"Each time I corrected an error, I noticed that the learners made another one and then I was not sure whether to correct them or not. I thought that they already had a lot to deal with, like focusing on the task, structuring, and typing their sentences, so I just did not want them to feel overwhelmed."

Given the fact that they were interacting with complete strangers, NSs/STs seemed to appreciate the idea of beginning with the chatting sessions before moving to the oral mode. They believed that this gave them some time to get to know each other and develop a bit of a rapport before they started having the oral conversations, where they had to be more careful about turn taking and where the tasks became a little more personal. Anna commented:

"I think starting with text-chat was actually helpful, especially for my partner, because the mode is a bit more of a protective base; it's safe, a bit more protected, a bit more anonymous maybe, and there is less exposure than voice chat."

5.3.2.2 <u>Perceptions towards the tasks</u>

Both NSs and STs found the tasks fun and useful. They believed that there was an interesting mixture, and that the tasks were pretty diverse in terms of what they had to do. NSs felt a bit worried though as some of the tasks were quite personal and directly related to the learners' life. Rosy, for example, commented:

"It is a bit unusual in the tasks that talk about things like personal values and family and friends. Those were quite personal things and I am not very familiar with people from Syria, so I did not know how conservative they would be about revealing you know details of their life or their feelings and I did not know how that would be for them in terms of the task. But I felt I am fine with it and I felt both learners have really a good sense of humour and were quite open."

In addition, NSs/STs found that the picture-based tasks were a little bit limited in terms of the kind of language that they elicited and that the latter tasks were more challenging; i.e., the learners were focusing more on their thoughts and their response rather than trying to overly monitor their language, so maybe they were being more natural. Hence, NSs/STs seemed to prefer decision-making tasks over other task types as they stimulated more use of the language and led to more authentic conversation. Nevertheless, they reported that they could see the logic in starting with simpler tasks and making them more complex as the sessions were progressing.

5.3.2.3 <u>General Comments:</u>

Apart from Rosy, who offered explicit feedback more often, especially during the text-chat, NSs/STs seemed to follow the same approach when correcting learners' errors which was mostly implicit. Rosy explained:

"I did not want to talk down to them because they are adults, but I thought it's useful to highlight or raise their awareness of certain errors/points. I tried to be pretty quick so that it was still fresh in their mind and they could immediately draw a connection."

All the other participants, who provided mainly implicit feedback, reported that they felt uncomfortable interrupting learners and that they did not know the learners very well or how they preferred to be corrected. Sally, one of the STs, said:

"This was the tricky part; I just wasn't sure what is the best way to approach these errors as I don't know the learners before. I chose not to be direct in my correction and like of course not to correct all the errors as they might feel discouraged and stop speaking/writing. I think the point here is for them to talk and not to be over-corrected so it's fine as long as I am getting the idea."

NSs/STs were also asked about their feelings as they were interacting online with people they had not met before and how well they thought they worked together. They all emphasised that there was no discomfort on their part, and they were not put off by that as they are constantly used to being with a new group of people. However, Sally also pointed to the fact that like any learning/teaching context, some learners were more approachable and more motivated to learn than others, and in the case of the former, this made it easier for both interlocutors to build trust and rapport throughout the sessions. She said: "I believe that no matter what I did to help the learners improve and develop their language, if they did not really have the motivation to learn, I would get nowhere with them".

It was interesting though how some NSs/STs were able to tell when the students did not have as much to say about the topic and where they were getting a little impatient, and hence the former tried to mitigate any sort of frustration and offered more help or more language as they deemed necessary. Yet, it appeared that NSs/STs were not very sure how useful they were for the learners or how effective their feedback was. They traced this back to several factors: firstly, they said that the sessions were very short and spread apart; secondly, the tasks were different each week and thus there was not any recycling of the language; and finally, they were not able to get any feedback from the students on whether they learned new expressions, found this helpful, and were more aware of their errors as a result of the sessions. Rosy commented:

"I felt a little bit like I was sort of going through this blindly, so I think having a feedback process [from the students] would make it stronger for me, from a teacher's perspective".

Nevertheless, during the oral chat, some NSs/STs noticed that the learners were expressing themselves more and producing longer/more complex utterances compared with the text-chat. Melanie said:

"Obviously during the text chat, the learners were using very simple and short sentences and not really elaborating. I was often asking them, 'Oh, can you say a bit more about that?' and they did not write as much whereas when we were talking, I felt like they were giving me more, which I didn't expect actually."

They were unsure, however, whether that was due to the learners getting more confident, because they were progressing and they were gaining more control over their language, or perhaps a mix of both.

Summing up, it seems that NSs/STs had positive perceptions towards the online experience, yet they believed that having more training could have made them more comfortable about addressing learners' errors and perhaps have made them more useful for their partners. Reflecting back on the whole thing, Anna commented:

"something that I did not expect, I found teaching online a bit more difficult than I thought; things like how do you correct errors? I haven't really thought about it, and actually, I don't really have an answer to that. It's something that I guess if I am gonna do the project again, I will probably do a bit more research about it and kind of think slightly more in depth about how that kind of thing could happen".

Similarly, Melanie mentioned that: "I think if I had the opportunity to have like some sort of training and get like a few more tips on how like feedbacking in a chat then I think I could be more effective in my role". Kate, on the other hand, believed that: "online teaching would be

good for group work and like facilitating the conversation between two people. I think that would be really good".

6. <u>Conclusion</u>

The overarching aim of this study was twofold: 1) first, to investigate whether manipulating task complexity along with a pre-planning/no planning condition during the text chat/voice chat sessions had an impact on learners' L2 development; and second, to examine whether the oral and written proficiency of intermediate level Syrian learners improved as a result of their longitudinal online interaction with more proficient L2 users. This study also examined the strategies used by the Syrian learners when planning for different task types across different modes of online interaction (text vs. voice chat); identified the factors that impacted learners' L2 development across different modes of online interaction (text vs. voice chat); and lastly inspected the attitudes and perceptions of the participants towards the online experience.

I begin this chapter by providing a synthesis of the main findings and explaining how they take forward current understanding in terms of theory, methodology, and pedagogy. Then, I discuss the limitations and drawbacks of the present research. Finally, the last section describes areas for future research.

6.1 Theoretical implications:

Theoretically, the current research is significant as it aims to contribute to the body of knowledge in the CAF literature on two different modes of SCMC. That is, the experimental work presented here provides one of the first investigations into the combined effects of manipulating task complexity along with pre-planning/no planning conditions via means of text-based and voice chat on learners' subsequent written L2 production across immediate and delayed post-tests. Data analysis of learners' output throughout the tests revealed that:

- the treatment offered to the experimental group (as opposed to the control group) during the online interaction, including the provision of planning time prior to task

performance as well as guided instructions to employ particular linguistic forms, led to greater syntactic complexity and more syntactically/lexically varied output of learners' written narratives during the immediate and delayed tests.

▪ The treatment group also showed noticeable progress (over the control group) regarding the number of words they produced per test (fluency). However, no significant difference was found between the groups in terms of the accuracy of learners' production (see **table 52**).

Hence, it appeared that complexity and fluency were often favoured at the expense of accuracy during the written tests as the tasks were getting more complex. Such findings could be explained with reference to Skehan's (1998; 2009) trade-off theory, which predicts that cognitively demanding tasks will hinder learners' ability to attend to multiple features of their L2 production at once, leading to trade-off effects between complexity and accuracy/accuracy and fluency. That is to say, learners' focus on complexity and fluency possibly compromised their performance with reference to accuracy. Additionally, this seems in accordance with Michel's (2017) argument that offering learners time to pre-plan their tasks "is likely to increase complexity and fluency because L2 speakers can conceptualize their performance beforehand" (p.60). Perhaps this was also transferred to learners' subsequent performance during the tests. Thus, the CAF results of the written tests in this study appeared to contradict Robinson's (2005) argument that pre-task planning serves to simplify the task, leading to greater accuracy as it assists automatic access to the stored language.

Table 52: A Summary of the significant results for each CAF measure throughout the written and oral tests

CAF Measures	Written					Oral				
	pre	*post 1*	*delayed 1*	*post2*	*delayed 2*	*pre*	*post1*	*Delaye d1*	*post 2*	*Delayed 2*
Cm1		✓	✓	✓	✓		✓	✓	✓	✓
Cm2		✓	✓	✓	✓		✓	✓	✓	✓
Cm3				✓						
Cm4					✓					
Cm5		✓	✓					✓		
Cm6				✓	✓					
Am1							✓			✓
Am2							✓			✓
Fm1 (written chat only)	✓		✓	✓		-	-	-	-	-
Fm1	-	-	-	-	-	✓		✓		✓
Fm2	-	-	-	-	-					
Fm3	-	-	-	-	-					✓
Fm4	-	-	-	-	-			✓	✓	✓

Findings from learners' speaking tests also revealed that:

- the experimental group benefitted from the treatment they received during the online sessions, producing language that was more syntactically complex and occasionally more lexically varied than the control group; not much change was visible for the other complexity measures, though.

- Unlike findings from the learners' written narratives (during the post/delayed tests), independent samples t-tests showed that the treatment group were more accurate in their oral narrations than the control group.

▪ Concerning the fluency of learners' production (measured by the number of filled/silent pauses, number of meaningful syllables, and finally number of repairs), the treatment group also achieved more progress compared with the control group.

Although such results can be taken to support Robinson's (2001) cognition hypothesis since the experimental group managed to achieve progress across the three CAF dimensions, it is apparent that the learners were not attending to all the CAF subdimensions at the same level (i.e., significant improvements were not always achieved simultaneously for each subcomponent per test) (see **table 52**). Thus, once again, and despite the treatment they received during the sessions, the experimental group did focus their attention on some CAF constructs/subconstructs to the detriment of others throughout the subsequent spoken tests.

6.2. Methodological implications

As opposed to the majority of previous CAF research which relied on short-treatment duration (e.g., Adams et al., 2015; Ellis & Yuan, 2004), the longitudinal design of the current study seemed to have positive effects on learners' overall L2 proficiency, leading to significant methodological implications. First, this design was found to be promising as it not only offered learners multiple authentic opportunities to practise written/spoken English online with more proficient language users, but also enabled me to utilise a variety of task types and observe the participants' online behaviour over a period of fifteen consecutive weeks. Second, the pre-, post-, and delayed tests employed in this study were essential to track learners' progress per group and detect any changes in their L2 proficiency (as indexed by global and specific CAF measures). Greater short-term/long-term gains were uncovered for the treatment group when their CAF scores in the pre-test and those of the following tests were compared. Specifically, the treatment group produced more syntactically complex output during the subsequent written

and spoken tests, yet no significant progress was identified for the other complexity measures in any of the tests. The findings also revealed that the affordances of the text-based chat mode, the permanency of the online messages and the saliency of NS'/STs' feedback as projected on learners' screens, exerted a greater positive influence on the accuracy of learners' narratives in the written but not in the spoken tests. This was expected given the fact that learners' production in the latter was more instant and hence there was less time to think about the quality of their language. Finally, when we examined the fluency of learners' production throughout the tests, results were mixed. Significant gains, however, were detected following the voice chat sessions in terms of the number of words/syllables and repair moves produced per test, which indicated a positive influence of these sessions on learners' fluency. Accordingly, the utilisation of written and oral online tools for L2 practice in the present work generates fresh insights into how the unique features of each modality could be best exploited to shape opportunities for L2 learning and L2 development. However, adding a visual element to learners' interaction could have made the learning process more efficient and more like F2F contexts. Besides, it seems that learners' exposure to the targeted language in the current study was not enough to impact all the selected CAF measures (especially because there was no other way for them to use the language in their context). Or maybe the CAF measures used to operationalise L2 proficiency were not the most appropriate and well-controlled to fit the aims of this study. Therefore, making the practice sessions more frequent (two or three per week), and/or integrating more fine-grained measures could perhaps have resulted in more gains in learners' L2 proficiency (Michel, 2017).

Another methodological implication pertained to the incorporation of stimulated recall interviews in this study which were conducted with the SLs following each chat session. Valuable insights were uncovered as to why the learners behaved in a particular way, and/or achieved more/less progress throughout the course of the study. Answers to these questions provided information that could not have been collected otherwise. The interviews with the chosen four case studies, for example, revealed the following:

- There were differences amongst the learners in terms of their planning strategies either before or during the online sessions. Majd and Lara, for example, gradually chose to attend to what Ortega (2005) terms 'form-in-meaning' where they tended to "pay attention to the inextricable relationship between form and meaning, simultaneously holding in long-term memory considerations regarding the message to be conveyed and the essential formal resources to convey it" (p. 106). Hence, in accordance with Ortega's (2005) findings, the planning time/guided instructions provided prior to the online sessions helped some learners strike a balance between focussing on the communicative demands of the targeted tasks as well as the language needed to complete these tasks. The other two learners (Ameen and Rima), however, appeared to have a natural inclination to focus on meaning when writing/speaking in their L2, without being overly concerned about the formal aspects of their production.

- There were a few factors that substantially influenced learners' planning decisions as well as how they perceived the benefits of the planning opportunities. Firstly, as the tasks were getting more complex, learners' attentional resources during the preparation time were primarily directed towards meaning. Secondly, as soon as the audio chat sessions started, some learners were found to be more appreciative of the ten minutes provided beforehand to prepare for their online performance since, unlike text chat, the

279

conversation in the oral mode was more instantaneous and thus less time was available to contemplate the content of the task. Finally, the L1 status of the more proficient language user, being a NS, influenced learners' planning choices as they aimed to produce output which was meaningful and well-structured.

It is worth emphasizing that apart from the planning opportunities, type/content of the tasks, mode of interaction, and learners' production of more/fewer complex structures, errors and pauses during the online sessions were also influenced by the amount/type of feedback delivered by NSs/STs, rapport between the learners and their partners, and/or learners' motivation/anxiety levels. These, in turn, could have also impacted learners' production during the subsequent tests (i.e., their CAF gains) and affected the consistency of the findings.

6.3 Pedagogic Implications

The findings of the current study lead to important pedagogical implications for language learning and teaching. Almost all the participants appreciated this experience, enjoyed the online sessions and saw chatting as both fun and beneficial for language learning. They also talked about the different features and affordances of SCMC that might help them improve their English. Accordingly, SCMC that encourages one-to-one interaction (written/spoken) between L2 learners and NSs or more experienced L2 interlocutors constitutes an optimal environment for interlanguage development. Both written and spoken interaction can be associated with pedagogical benefits. Taking written interaction first, the saliency of L2 production and the slow pace of the written interaction gave the learners the opportunity to review and edit their messages before sending them. In contrast, the speed of interaction in the voice chat mode more closely resembles F2F conversation and allows far more L2 production in a shorter time. Hence language teachers can take advantage of the distinctive affordances of these two modes to scaffold the learners to produce more complex, accurate, and fluent output.

Secondly, via the incorporation of SCMC tools into course syllabi, language instructors can establish a supportive learning atmosphere, through which L2 learners communicate together to achieve meaningful objectives beyond the confines of the classroom walls, and hence enhance their communication competences.

An analysis of the aggregated data (i.e., learners' chat logs, recall interviews) also showed that providing learners with external guidance to consider the integration of forms that might be necessary/useful for task completion would automatically orient learners' attention towards these forms. Hence, another implication of this study for syllabus designers and teachers is associated with the guided planning treatment which not only helps reduce the cognitive load learners might experience while performing increasingly complex tasks and directs their attentional resources to multiple aspects of the language in predictable ways (Skehan, 1998), but also has the added advantage of drawing learners' attention to incorporate forms which are either known for being complex such as relative clauses (Ortega, 2005; Thompson, 2014), and/or have been frequently considered low in their saliency (e.g., articles) (Ortega, 1999).

As guided planning entails drawing learners' attention to language forms (i.e., form and/or content), the question for SLA researchers as well as teachers remains as to which type of guided planning would benefit learners' proficiency the most (Mochizuki & Ortega, 2008).

Likewise, task-based collaborative online chat in SCMC can be used to draw learners' attention to linguistic forms through a careful design of communicative activities. In other words, integrating tasks which differ in their cognitive demands and sequencing tasks from simple to complex (based on Robinson's (2003) Cognition Hypothesis) seems to provide equal opportunities for learners to attend to different aspects of a language and consequently lead to optimal gains in L2 development. Hence, the implication for syllabus designers and teachers integrating this technological tool into their teaching approaches is that cognitive complexity is a reliable and robust criterion for choosing, designing, and sequencing of pedagogical tasks.

In other words, the complexity of a particular task can be manipulated in online discourse with the aim of matching students' progression as well as their proficiency levels. According to Ellis (2003), "presenting and practising features learners have failed to use correctly in production may not result in their acquisition if the learners are not developmentally ready to acquire them" (p. 30). The manipulation of task complexity can also help optimise learners' opportunities to notice the gaps they have in the targeted language and eventually lead to interlanguage development.

Finally, chat logs give teachers instant access to learners' output data, and hence these logs might possibly be a beneficial source for planning lessons in communicative L2 classrooms as they provide insights into learner interlanguage development. Simply put, evaluating chat logs allows the advantage of witnessing the processes which the learners go through while seeking to achieve certain communicative goals. In addition, language learners might benefit from thoroughly scrutinizing their chat logs as they contain permanent records of their language progression.

6.4 Limitations of the current study

Due to the small number of participants, the results in the current study were cautiously interpreted, restricting broad generalization with regard to the combined effects of manipulating task complexity and planning conditions on L2 development. Additionally, the intermittent nature of Internet coverage in Syria and the lack of access to a keystroke logger did not allow me to monitor and record each keystroke typed by the SLs; and therefore, it was not possible to examine the impact of the online treatment received by the planning group on the fluency of their written production.

Another limitation of this study is associated with learners' proficiency; all the learners recruited in the present study were of an intermediate-level, and hence we do not know whether learners of higher/lower proficiency level would act differently when working under similar conditions. Thus, more research is needed on learners of different language proficiency levels. A further limitation relates to the context where the current study was carried out. That is, data collection was conducted in an experimental, non-classroom based setting and varied tasks were employed without associating their use with the teaching of an ongoing programme of instruction. Therefore, the significant progress achieved by the treatment group, in particular, with regard to the majority of CAF measures is disconnected from the context in which the teaching/learning process took place. In addition, the control group in the current study also showed some improvement in their L2 performance. As the data later showed, different external factors affected the results of this study, amongst which were learners' motivation level and rapport between the interlocutors. Some learners, regardless of the group they were assigned to, were desperate to learn English as they wanted to move abroad to find a job or pursue their studies. Hence, they were helping themselves, taking risks, asking questions, and requesting feedback from their partner even when they were not instructed to do so. Extensive pedagogical contextualised research is needed in order to evaluate the effectiveness of manipulating task complexity along with pre-planning/no-planning conditions on L2 development (Bygate et al. (2009).

One more limitation with the findings of this study is the highly varied reactions of the STs and NSs as regards correction despite the explicit instructions they were given to try to address their partners' errors during the online sessions. In hindsight, I think that one of the things I could have done was including a much longer, more rigorous period of training for the STs/NSs, in order to get a far more uniform approach to correction. Piloting certainly helped, as it resulted in me realizing that I needed to make the instructions about correcting students more explicit

for teachers, and in fact, this worked to some extent. But there was still a wide spectrum of correction practices in evidence, and it would have needed perhaps one week's training and practising for the teachers in giving feedback.

6.5 Suggestions for Future Research

The findings of the present study have synthesised several possibilities for future research. These are grouped under different subheadings in the following paragraphs:

6.5.1 task type and complexity

Though the tasks were sequenced in the current study based on Robinson's (2005) criteria for establishing task complexity, variations in individual differences could have affected how learners viewed these tasks, i.e., simple or complex (e.g., learners' knowledge about and understanding of the task's content). Therefore, and following Allaw & McDonough's (2019) suggestion, further research is needed to provide a more robust empirical basis for sequencing tasks by exploring how the manipulation of specific task features (along both resource-directing and resource-dispersing variables) can result in series of simple-to-complex tasks to be then used in F2F contexts and/or online environments. This will provide language teachers and those responsible for curriculum designers with the information needed to come up with informed decisions on how to design and implement tasks in a variety of instructional settings; in ways that trigger learners to attend to multiple aspects of their L2 production in order to achieve the goal of balanced development. In addition, the present study investigated the role of guided planning and task complexity on the complexity, accuracy and fluency of learners' written/spoken production. Thus, an important avenue for future research would be to investigate how task grading may have affected SLs' writing/speaking processes (e.g., editing and monitoring). Future studies should also ensure that all the implemented tasks are imposing an acceptable degree of challenge for their participants.

Based on the findings of the current study that showed how a change in task type induced learners to attend to different aspects of their production, more research should be done to closely examine whether learners' focus for the same task type would change based on the mode of task delivery. It would also be interesting to examine the effects of other task types on L2 performance (e.g., problem-solving, dictogloss task).

6.5.2 Types of interlocutors

The current study recruited learners with an intermediate proficiency level, and the findings showed that the treatment received by the planning group did have some positive effects on learners' performance and resulted in significant gains in terms of CAF. Hence, it would be worthwhile investigating whether or not manipulating task complexity via SCMC can promote L2 development with lower/higher proficiency level learners. Future research may also scrutinise whether pairing learners with different proficiency levels has any impact on their overall performance, particularly when working with an English teacher and/or NS is not feasible.

It is worth mentioning that to the best of my knowledge, this was the only study that examined the combined effects of manipulating task complexity along planning conditions during two different modes of online interaction with more proficient L2 speakers (i.e., NSs/STs). Previous planning studies in SCMC were conducted primarily in text chat, where learners had to either narrate a story to the researcher or work with another learner who has the same proficiency level (e.g., Hsu, 2012, 2015; Ziegler, 2018). NSs/STs in the current study were instructed to trigger learners' attention to use particular linguistic forms (i.e., relative clauses), reformulate learners' utterances, and provide feedback on learners' errors. However, it was beyond the scope of the current study to investigate whether the learners experience comparable opportunities for language learning when interacting with a NS or a ST, and whether there are

any differences in the feedback provided by each. These are definitely issues that deserve further examination.

6.5.3 Types of online interaction

There has been a massive increase in the use of technology for teaching/learning purposes especially in the last year, underlining the demand for methodologically well-grounded, relevant research to inform pedagogical practice. Much of the research on the effectiveness of preparation time, task-based instructions, and feedback opportunities on L2 proficiency has been conducted in written text-chat (e.g., Adams et al. 2015; Hsu, 2012; Sauro & Smith, 2010; Ziegler, 2018). Hence, based on the findings of the current study that showed some improvements in learners' production while interacting with their partners via means of voice chat, future studies need to cast more light on the affordances of other modes of SCMC (voice and video chat) that bear a strong resemblance to F2F discourse. It would also be interesting to examine the immediate and sustained effectiveness of target form knowledge (i.e., forms that have been corrected) as well as the types of feedback delivered via video chat on learners' L2 development. This would provide further support for the potential of different modes of SCMC for language learning.

6.5.4 Types of CAF measures

Due to the lack of longitudinal studies that examine learners' L2 performance (operationalised by different CAF measures) while working on dialogic rather than monologic tasks via SCMC, there is a pressing need for future SCMC research that scrutinises the development of other language forms which are also recognised for their difficulty in L2 production in a particular context. According to Bulte & Housen (2012), the complexity of these linguistic features may be influenced by factors such as the learner's L1 background.

Additionally, the lack of access to a keystroke logger in the present study did not allow me to monitor and record each keystroke typed by the SLs which did not make it possible to utilise fluency measures for the written sessions/tests. Further research that employs this tool is needed, therefore, to measure the fluency of learners' written production and to get more insights into the cognitive process the learners involve in while typing (i.e., editing, disfluencies, monitoring). Future research might also include a variety of general and specific CAF measures to allow broader comparisons with previous research.

6.5.5 Types of programme

A very important and thriving area for future research has to do with the impact of manipulating task complexity via SCMC within university online English course programmes, where learners have more practice opportunities and more frequent exposure to the language (compared to the current study). Although this may seem like a bit of a challenge, it is hoped that the significant CAF findings uncovered in the present study (which was conducted in an experimental setting) will inspire future inspections into the role of guided instructions, planning time, and task complexity, and that future research is warranted to scrutinise how effective the combination of these three variables could be within a university online English course programme. Apart from the selection of tasks that meet learners' needs, practice/training sessions should be implemented before the programmes commence, so that the English teachers can get some experience on how to feedback on learners' errors during the online interaction. Furthermore, there should be some practice sessions for the learners as well to ensure their familiarity with note-taking strategy for different task types and track their behaviour in the planning time. This kind of training would influence how learners approach and plan for each task type, so that they can make the best out of the planning time provided before the actual task performance.

References

Abrams, Z. I. (2003). The effect of synchronous and asynchronous CMC on oral performance in German. *Modern Language Journal*, 87(2), 157–167.

Adams, R., Alwi, N., Newton, J. (2015). Task complexity effects on the complexity and accuracy of writing via text chat. *Journal of Second Language Writing,* 29: 64-81.

Ahmadian, MJ., Tavakoli, M., and Dastjerdi, H. (2015). The combined effects of online planning and task structure on complexity, accuracy, and fluency of L2 speech. *Language Learning Journal*, 43(1), 41-56.

Ahmadian, M.J. (2012). The effects of guided careful online planning on complexity, accuracy, and fluency in intermediate EFL learners' oral production: The case of English articles. *Language Teaching Research*, 16(1), 129–49.

Ahmadian, M.J., & Tavakoli, M. (2010). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research* 15(1), 35–59.

Allaw, E., & McDonough, K. (2019). The effect of task sequencing on second language written lexical complexity, accuracy, and fluency. *System*, 85: 1-12.

Alwi, N., Adams, R., & Newton, J. (2012). Writing to learn via text chat: Task implementation and focus on form. *Journal of Second Language Writing*, 21, 23-39.

Asgarikia, P. (2014). The Effects of Task Type, Strategic Planning and No Planning on Written Performance of Iranian Intermediate EFL Learners. *Procedia - Social and Behavioral Sciences*, 8: 276 – 285.

Aydin, Z., & Yildiz, S. (2014). Using wikis to promote collaborative EFL writing. *Language Learning & Technology*, 18(1), 160–180.

Baralt, M., & Leow, R. P. (2016). Uptake, task complexity, and L2 development in SLA: An online perspective. In R. P. Leow, L. Cerezo & M. Baralt (Eds.). *A psycholinguistic approach to technology and language learning* (pp. 3-22). Berlin; Boston: De Gruyter Mouton.

Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition*, 35, 689–725.

Beauvois, M. H. (1998) Conversations in slow motion: Computer-mediated communication in the foreign language classroom. *Canadian Modern Language Review*, 54(2), 198–217.

Bikowski, D., & Vithanage, R. (2016). Effects of Web-based Collaborative Writing on Individual L2 Writing Development. *Language Learning & Technology,* 20(1), 79-99.

Blake, R., Wilson, N., Cetto, M., & Pardo-Ballester, C. (2008). Measuring oral proficiency in distance, face-to-face, and blended classrooms. *Language Learning & Technology*, 12(3), 114-127.

Blake, R., & Zyzik, C. (2003). Who's helping whom? Learner heritage-speakers' networked discussions in Spanish. *Applied Linguistics*, 24(4), 519-544.

Blake, R. (2000). Computer mediated communication: A window on L2 Spanish interlanguage. *Language Learning & Technology*, 4(1), 120-136.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org/

Bower, J., Kawaguchi, S., (2011). Negotiation of meaning and corrective feedback in Japanese/English e-tandem. *Language Learning & Technology,* 15(1), 41-71.

Bulte, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency* (pp. 21-47). Amsterdam: John Benjamins.

Bueno-Alastuey, M.C. (2013). Interactional feedback in synchronous voice-based computer mediated communication: Effect of dyad. *System*. 41, 543–559.

Bueno-Alastuey, M. C. (2011). Perceived benefits and drawbacks of synchronous voice-based computer-mediated communication in the foreign language classroom. *Computer Assisted Language Learning*, 24(5), 419-432.

Bui, G., & Skehan, P. (2018). Complexity, Accuracy, and Fluency. In: J. Liontas (Ed.), *TESOL Encyclopedia of English language teaching* (pp. 1-7). Hoboken, NJ: Wiley-Blackwell.

Bygate, M., Skehan, P., & Swain, M. (2001). *Researching Pedagogic Tasks. Second Language Learning and Testing*. Harlow: Longman.

Canals, L. (2020). The effects of virtual exchanges on oral skills and motivation. *Language Learning & Technology*, 24 (3), 103-119.

Chen, W. C., & Eslami, Z. R. (2013). Focus on form in live chats. *Educational Technology & Society*, 16(1), 147-158.

Cho, H. (2011). Negotiation of meaning in synchronous computer-mediated communication in relation to task types (Unpublished doctoral dissertation). Indiana University, Bloomington, IN.

Chun, D. M. (1994). Using computer networking to facilitate the acquisition of interactive competence. *System*, 22 (1), 17-31.

Creswell, J.W., & Clark, V.L.P. (2017). *Designing and conducting mixed methods research*. Sage publications.

Crookes, G. (1989). 'Planning and interlanguage variation'. *Studies in Second Language Acquisition*, 11: 367–83.

De Bot, K. (1992). A bilingual production model: Levelt's 'Speaking' model adapted. *Applied Linguistics*, 13 (1), 1-24.

DeKeyser, R. M. (2005). What makes learning second language grammar difficult? A review of issues. *Language Learning*, 55(1), 1-25.

de Jong, N. H., Steinel, M., Florijn, A., Schoonen, R., & Hulstijn, J. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, 34(1), 5–34.

Dörnyei, Z. (2007) *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies.* Oxford: Oxford University Press.

Doughty, C., & Pica, T. (1986). Information-gap tasks: Do they facilitate SLA? *TESOL Quarterly*, 20(2), 305-325.

Ellis, R. (2010) Second language acquisition research and language-teaching materials. In N. Harwood (ed), *English Language Teaching Materials: Theory and Practice* (pp. 33-57). CUP.

Ellis, R. (2009a). The effects of three types of task planning on the Fluency, Complexity, and Accuracy in L2 Oral Production. *Applied Linguistics*, 30(4), 474–509.

Ellis, R., (2005). Planning and task-base performance: theory and research. In: Ellis, R. (Ed.), *Planning in Task-Based Performance* (pp. 1-36)*.* John Benjamins, Amsterdam.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Ellis, R., & Barkhuizen, G. (2005). *Analysing Learner Language*. Oxford: Oxford University Press.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26: 59–84.

Elola, I., & Oskoz, A. (2010). Collaborative writing: Fostering foreign language and writing conventions development. *Language Learning & Technology*, 14(3), 51–71.

Ericsson, K. A. (2002). Towards a Procedure for Eliciting Verbal Expression of Non-verbal Experience without Reactivity: Interpreting the Verbal Overshadowing Effect within the Theoretical Framework for Protocol Analysis. *Applied cognitive psychology*, 16: 981–987.

Eslami, Z. R., & Kung, W. T. (2016). Focus-on-form and EFL learners' language development in synchronous computer-mediated communication: task-based interactions. *The Language Learning Journal*, 44(4), 401-417.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons. *Applied Linguistics*, 21(3), 354-375.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18: 299-323.

Gass, S. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Erlbaum.

Gass, S., & Mackey, A. (2017). *Stimulated recall methodology in applied linguistics and L2 research*. London: Routledge.

Gass, S. M., & Varonis, E. (1994). Input, interaction, and second language production. *Studies in Second Language Acquisition*, 16 (3), 283-302.

Gilabert, R. (2007b). The simultaneous manipulation of task complexity along planning time and [+- Here-and-Now]. In M.D.P. Garcia-Mayo (Eds.), *Investigating tasks in formal language learning* (pp. 44-68). Clevedon, UK: Multilingual Matters.

Gonzalez-Lloret, M. (2017). Technology for task-based language teaching. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 234–247). Hoboken, NJ:Wiley Blackwell.

Granena, G. (2016). Individual versus interactive task-based performance through voice-based computer-mediated communication. *Language Learning & Technology*, 20(3), 40-59.

Gurzynski-Weiss, L., Al Khalil, M., Baralt, M., & Leow, R.P. (2015). Levels of awareness in relation to type of recast and type of linguistic item in synchronous computer-mediated communication: A concurrent investigation. In R. P. Leow, L. Cerezo, & M. Baralt (Eds.), *A psycholinguistic approach to technology and language learning* (pp. 151-170). Berlin: De Gruyter Mouton.

Gurzynski-Weiss, L., & Baralt, M. (2014). Exploring learner perception and use of task-based interactional feedback in FTF and CMC modes. *Studies in Second Language Acquisition*, 36(1), 1-37.

Hammerly, H. (1991). *Fluency and Accuracy: toward Balance in Language Teaching and Learning*. Clevedon: Multilingual Matters.

Hampel, R., & Hauck, M. (2004). Towards an effective use of audio conferencing in distance language courses. *Language Learning and Technology*, 8(1), 66-82.

Handley, Z. L. (2014). Constructing an evidence-base for future CALL design with 'engineering power': The need for more basic research and instrumental replication. *EUROCALL Review*.

Henderson, C. (2021). The effect of feedback timing on L2 Spanish vocabulary acquisition in synchronous computer-mediated communication. *Language Teaching Research,* 25(2), 185-208.

Housen, A., Kuiken, F., & Vedder, I. (2012), Complexity, accuracy and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 Performance and Proficiency* (pp. 1-21). Amsterdam: John Benjamins.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461-473.

Hsu, H. (2017). The Effect of Task Planning on L2 Performance and L2 Development in Text-Based Synchronous Computer-Mediated Communication. *Applied Linguistics*, 38(3), 359-385.

Hsu, H. (2012). Investigating the Effects of Planning on L2 Text Chat Performance. *CALICO*, 29(4), 619-638.

Jepson, K. (2005). Conversations and negotiated in teraction in text and voice chat rooms. *Language Learning and Technology*, 9(3), 79-98. Retrieved September 15, 2007, from http://llt.msu.edu/vol9num3/jepson/default.html

Kawauchi, C. (2005). The effects of strategic planning on the oral narratives of learners with low and high intermediate L2 proficiency. In R. Ellis (Eds.), *Planning and Task Performance in a Second Language* (pp. 143-165). Amsterdam: John Benjamins.

Kern, R. G. (1995). Restructuring classroom interaction with networked computers: Effects on quantity and characteristics of language production. *The Modern Language Journal*, 79: 457–476.

Ko, C. (2012). Can synchronous computer-mediated communication (CMC) help beginning-level foreign language learners speak? *Computer Assisted Language Learning*, 25(3), 217-236.

Koizumi, R. (2012). Relationships Between Text Length and Lexical Diversity Measures: Can We Use Short Texts of Less than 100 Tokens? *Vocabulary Learning and Instruction*, 1(1), 60-69.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32, 145–164.

Kost, C. R. (2004) An investigation of the effects of synchronous computer-mediated communication (CMC) on interlanguage development in beginning learners of German: Accuracy, proficiency and communication strategies. University of Arizona, unpublished PhD.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing,* 20: 148–61.

Kuiken, F., Vedder, I., Housen, A., & De Clercq, B. (2019). Variation in syntactic complexity: Introduction. *International Journal of Applied Linguistics*, *29*(2), 161-170.

Kuiken, F., & Vedder, I. (2007). Cognitive task complexity and linguistic performance in French L2 writing. In M. P. Garcia-Mayo (ed.): *Investigating Tasks in Formal Language Learning.* Multilingual Matters.

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing,* 17: 48-60.

Lai, C., & Zhao, Y. (2006). Noticing and text-based chat. *Language Learning & Technology*, 10(3),102–120. Retrieved from http://llt.msu.edu/vol10num3/laizhao/default.html

Lambert, C., & Kormos, J. (2014). Complexity, accuracy, and fluency in task-based l2 research: Toward more developmentally based measures of second language acquisition. *Applied Linguistics*, 35(5), 607–614.

Larsen-Freeman, D. (2006). The Emergence of Complexity, Fluency, and Accuracy in the Oral and Written Production of Five Chinese Learners of English. *Applied Linguistics,* 27(4), 590-619.

Lee, C. and Polio, J. (2017). Written language learning. In S. Loewen and M. Sato (eds), *The Routledge Handbook of Instructed Second Language Acquisition*, 299–318. New York: Routledge.

Lee, L. (2008). Focus-on-form through collaborative scaffolding in expert-to-novice online interaction. *Language Learning & Technology*, 12(3), 53-72.

Lennon, P. (2000). The Lexical Element in Spoken Second Language Fluency. In H. Riggenbach (Eds.), *Perspectives on Fluency* (pp. 25-42). Ann Arbor: University of Michigan Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Lim, S. (2002). The self-confrontation interview: towards an enhanced understanding of human factors in web-based interaction form improved website usability. J. Electron. *Commer. Res,* 3: 162–173

Lin, H. (2015). Computer-mediated communication (CMC) in L2 oral proficiency development: A meta-analysis. *ReCALL*, 27: 261-287.

Lin, W.-C., Huang, H.-T., & Liou, H.-C. (2013). The effects of text-based SCMC on SLA: A meta analysis. *Language Learning & Technology*, 17(2), 123–142.

Retrieved from http://llt.msu.edu/issues/june2013/linetal.pdf

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie, & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). New York: Academic Press.

Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In Crookes, G., & S.M. Gass (Eds.), *Task and language learning: Integrating theory and practice* (pp. 121–165). Clevedon: Multilingual Matters.

Lys, F. (2013). The development of advanced learner oral proficiency using iPads. *Language Learning & Technology*, 17(3), 94–116.

Mackey, A. (2012). *Input, interaction, and corrective feedback in L2 learning*. Oxford: Oxford University Press.

Mackey, A., & Gass, S. (2015). *Second Language Research: Methodology and Design* (2nd ed.). New York: Routledge.

McCarthy, P. M. (2005). An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD). Memphis, TN: University of Memphis. Unpublished doctoral dissertation.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20: 83-108.

Michel, M., Murakami, A., Alexopoulou, T., & Meurers, D. (2019). Effects of task type on morphosyntactic complexity across proficiency: evidence from a large learner corpus of A1 to C2 writings. *Instructed Second Language Acquisition,* 3(2), 124-152.

Michel, M. (2017). Complexity, accuracy, and fluency in L2 production. In S. Loewen, & S. Masatoshi (Eds.), *Routledge handbook of instructed second language acquisition* (pp. 50–68). New York: Routledge.

Mitchell, R., Myles, F., & Marsden, E. (2013). *Second language learning theories*. Abingdon: Routledge.

Mochizuki, N., & Ortega, L. (2008). Balancing communication and grammar in beginning level foreign language classrooms: A study of guided planning and relativization. *Language Teaching Research,* 12(1), 11–37.

Muñoz, C. (2006). The effects of age on foreign language learning: The BAF Project. In C. Muñoz (ed.), *Age and the rate of foreign language learning* (pp. 1-40.). Clevedon: Multilingual Matters.

Nation, P. (2015). Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language*, 27(1), 136-145.

Nisbett, R., & Wilson, T. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review,* 84(3), 231-259.

Norris, J., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: the case of complexity. *Applied Linguistics,* 30(4), 555-578.

Ong, J., & Zhang, L.J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing* 19: 218-233.

Ortega, L. (2005). What do learners plan? Learner-driven attention to form during pre-task planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 77–109). Amsterdam/Philadelphia: John Benjamins.

Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, 21: 109-148.

O'Rourke, B. (2005). Form-focused interaction in online tandem learning. *CALICO Journal*, 22(3), 433–466.

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, 31 (1), 117–134.

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30(4), 590-601.

Payne, S., & Whitney, P. (2002). Developing L2 oral proficiency through synchronous CMC: Output, working memory, and interlanguage development. *CALICO Journal*, 20(1), 7–32.

Pellettieri, J. (2000). Negotiation in cyberspace: The role of chatting in the development of grammatical competence. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching: Concepts and practice* (pp. 59–86). Cambridge: Cambridge University Press.

Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning*, 44: 429-527.

Plano Clark, V. L. & Ivankova, N. V. (2016). Mixed methods research. *A guide to the field*. Sage Publications.

Révész, A., Ekiert, M., & Torgersen, E. (2016). The effects of complexity, accuracy, and fluency on communicative adequacy in oral task performance. *Applied Linguistics*, 37(6), 828 848.

Révész, A. (2011). Focus on L2 Constructions, and Individual Differences: A Classroom-Based Study. *Modern Language Journal*, 162–181.

Richards, K. (2003). *Qualitative Inquiry in TESOL*. Basingstoke: Palgrave Macmillan.

Richards, B., & Malvern, D. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19: 85-104.

Robinson, P. (Ed.). (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. Amsterdam/Philadelphia: John Benjamins.

Robinson, P. (2010). Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In M. Putz, & L. Sicola (Eds.), *Inside the Learner's Mind: Cognitive Processing in Second Language Acquisition* (pp. 243-268). Amsterdam: John Benjamins.

Robinson, P., Cadierno, T., & Shirai, Y. (2009). Time and Motion: Measuring the Effects of the Conceptual Demands of Tasks on Second Language Speech Production. *Applied Linguistics,* 30(4), 533-554.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *IRAL*, 45(3), 193–213.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics*, 43: 1–32.

Robinson, P. (2003). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies*, 21(2), 45-105.

Robinson, P., (2001). Task complexity, task difficulty, and task production: exploring interactions in a componential framework. *Applied Linguistics*, 21(1), 27-57.

Robson, C. (2002). *Real World Research*. Oxford: Blackwell.

Rogers, J., & Révész, A. (2019). Experimental and quasi-experimental designs. In J. McKinley & H. Rose (1ˢᵗ ed), *The Routledge Handbook of Research Methods in Applied Linguistics* (pp. 133-143). London. Routledge.

Samuda, V., & Bygate, M. (2008). *Tasks in Second Language Learning*. Basingstoke: Palgrave Macmillan.

Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Eds.), *Planning and Task Performance in a Second Language* (pp. 111-143). Amsterdam: John Benjamins.

Satar, M., & Ozdener, N. (2008). The Effects of Synchronous CMC on Speaking Proficiency and Anxiety: Text versus Voice Chat. *The Modern Language Journal*, 92(4), 595-613.

Sauro, S. (2012). L2 performance in text-chat and spoken discourse. *System*, 40: 335–48.

Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning & Technology*, 13(1), 96–120.

Sauro, S., & Smith, B. (2010). Investigating L2 performance in text-chat. *Applied Linguistics*, 31(4), 554-577.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3 – 32). Cambridge: Cambridge University Press.

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

Shannon-Baker, P. (2015). But I wanted to appear happy: How using arts-informed and mixed methods approaches complicate qualitatively driven research on culture shock. *International Journal of Qualitative Methods*, 14(2). 34-52.

Shekary, M., & Tahririan, M. H. (2006). Negotiation of meaning and noticing in text-based online chat. *The Modern Language Journal*, 90(4), 557-573.

Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual.* New York: Palgrave Macmillan Press.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics,* 30(4), 510–32.

Skehan, P. (2003). Focus on form tasks and technology. *Computer Assisted Language Learning,* 16(5), 391–411.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford University Press.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, 17(1), 35–62.

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.) *Cognition and second language instruction* (pp. 183-205). Cambridge, UK: Cambridge University Press.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), 93-120.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1: 185–211.

Smagorinsky, P. (2008) The Method Section as Conceptual Epicenter in Constructing Social Science Research Reports. *Written Communication,* 25: 389-411.

Smagorinsky, P. (1989). The reliability and validity of protocol analysis. *Written communication* 6(4), 463-479.

Smith, B. (2008). Methodological hurdles in captures CMC data: The case of the missing self-repair. *Language Learning & Technology*, 12(1), 85– 103.

Smith, B. (2003). Computer-mediated negotiated interaction: An expanded model. *The Modern Language Journal*, 87(1), 38-54.

Sotillo, S. (2005). Corrective Feedback via Instant Messenger Learning Activities in NS-NNS and NNS-NNS Dyads. *CALICO*, 22(3), 467-496.

Sotillo, S. M. (2000). Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning & Technology*, 4(1), 82–119.

Retrieved from http://llt.msu.edu/vol4num1/sotillo/default.html

Seyyedi, k., Ismail, A., Orang, M., Nejad, M. (2013). The Effect of Pre-Task Planning Time on L2 Learners' Narrative Writing Performance. *English Language Teaching*, 6(12), 1-10.

Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: exploring task effects. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 99–118). New York: Longman.

Tavakoli, H., Lotfi, A. R., & Biria, R. (2019). Effects of CALL-mediated TBLT on motivation for L2. Cogent Education, 6(1), 1-21. https://doi.org/10.1080/2331186X.2019.1580916

Tavakoli, P., Campbell, C., & Mccormack, J. (2016). Development of Speech Fluency Over a Short Period of Time: Effects of Pedagogic Intervention. *TESOL Quarterly,* 50(2), 447-471.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). Amsterdam: John Benjamins.

Thompson, C. (2014). Guided planning, task complexity and second language oral development. Unpublished PhD thesis.

Varonis, E., & Gass, S. M. (1985). Non-native/non-native conversations: A model for negotiation of meaning. *Applied Linguistics*, 6 (1), 71-90.

Warschauer, M. (1996). Comparing face-to-face and electronic communication in the second language classroom. *CALICO Journal*, 13 (2), 7-16.

Wendel, J. (1997). Planning and second language narrative production. Temple University, Japan. Unpublished doctoral dissertation.

Williams, J. (2005). Form-focused instruction. In E. Hinkel (ed.): *Handbook of Research in Second Language Teaching and Learning* (pp. 671-91). Lawrence Erlbaum Associates.

Willis, J. (1996). *A Framework for Task-Based Learning*. Harlow: Longman.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). Second language development in writing: Measures of fluency, accuracy, and complexity. Honolulu, HI: University of Hawaii, Second Language Teaching & Curriculum Center.

Yanguas, I., & Bergin, T. (2018). Focus on form in task-based L2 oral computer-mediated communication. *Language Learning & Technology*, 22(3), 65-81.

Yanguas, I. (2010). Oral Computer-Mediated Interaction between L2 Learners: It's about Time! *Language Learning & Technology*, 14(3), 72-93.

Yilmaz, Y. (2016). The linguistic environment, interaction and negative feedback. *Multilingualism and Second Language Acquisition*, 1, 1–86.

Yilmaz, Y., & Yuksel, D. (2011). Effects of communication mode and salience on recasts: A first exposure study. *Language Teaching Research*, 15, 457–477.

Yilmaz, Y., & Granena, G. (2010) The Effects of Task Type in Synchronous Computer-Mediated Communication. *ReCALL*, 22(1), 20-38.

Yin, R. (2014) Case study research: Design and methods. London: Sage

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, 24 (1), 1-27.

Yuksel, D., & Inan, B. (2014). The effects of communication mode on negotiation of meaning and its noticing. *ReCALL,* 26(3), 333-354.

Zalbidea, J. (2017). 'One task fits all'? The roles of task complexity, modality, and working memory capacity in L2 performance. *The Modern Language Journal*, *101*(2), 335-352.

Zeng, G., & Takatsuka, S. (2009). Text-based peer-peer collaborative dialogue in a computer-mediated learning environment in the EFL context. *System*, 37(3), 434–46.

Ziegler, N. (2018). Pre-task planning in L2 text-chat: Examining learners' process and performance. *Language Learning & Technology*, 22 (3), 193-213.

Appendices

Appendix A
University of Sheffield

Participation Consent Form

Title of Research Project: Task Manipulation of Planning Conditions: An Analysis of Syrian Online L2 Performance in Text-Based & Voice Chat

Name of Researcher: Shahla Adi
Participant Identification Number for this project: Please initial box

1. I confirm that I have read and understand the information sheet dated _____*[insert date]* explaining the above research project and I have had the opportunity to ask questions about the project.

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without there being any negative consequences. In addition, should I not wish to answer any particular question or questions, I am free to decline.

3. I understand that my responses will be kept strictly confidential.

4. I give permission for members of the research team to have access to my anonymised responses.

5. I understand that my name will not be linked with the research materials, and I will not be identified or identifiable in the report or reports that result from the research.

6. I agree for the data collected from me to be used in future research

7. I agree to take part in the above research project.

_____   _____   _____
Name of Participant   Date   Signature
(*or legal representative*)

__Shahla Adi_____   _____   _____
Name of person taking consent   Date   Signature

To be signed and dated in presence of the participant

_____Shahla Adi_____   _____   _____
 Lead Researcher   Date   Signature
To be signed and dated in presence of the participant

E-mail: sadi1@sheffield.ac.uk   Skype ID: sha096
Contact Number: UK mobile: +447780277866
Once this has been signed by all parties the participant should receive a copy of the signed and dated participant consent form, the letter/pre-written script/information sheet and any other written information provided to the participants. A copy of the signed and dated consent form should be placed in the project's main record (e.g. a site file), which must be kept in a secure location.

I. **Narrative Tasks used during the piloting:**
   1. **(adopted from Mochizuki & Ortega, 2008)**
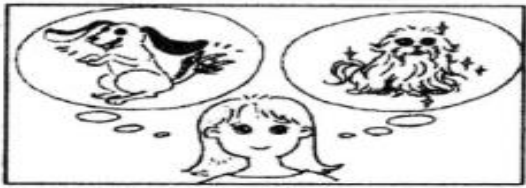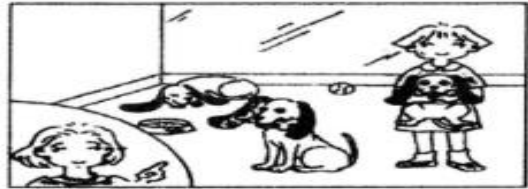
2.   (adopted from Thompson, 2014)
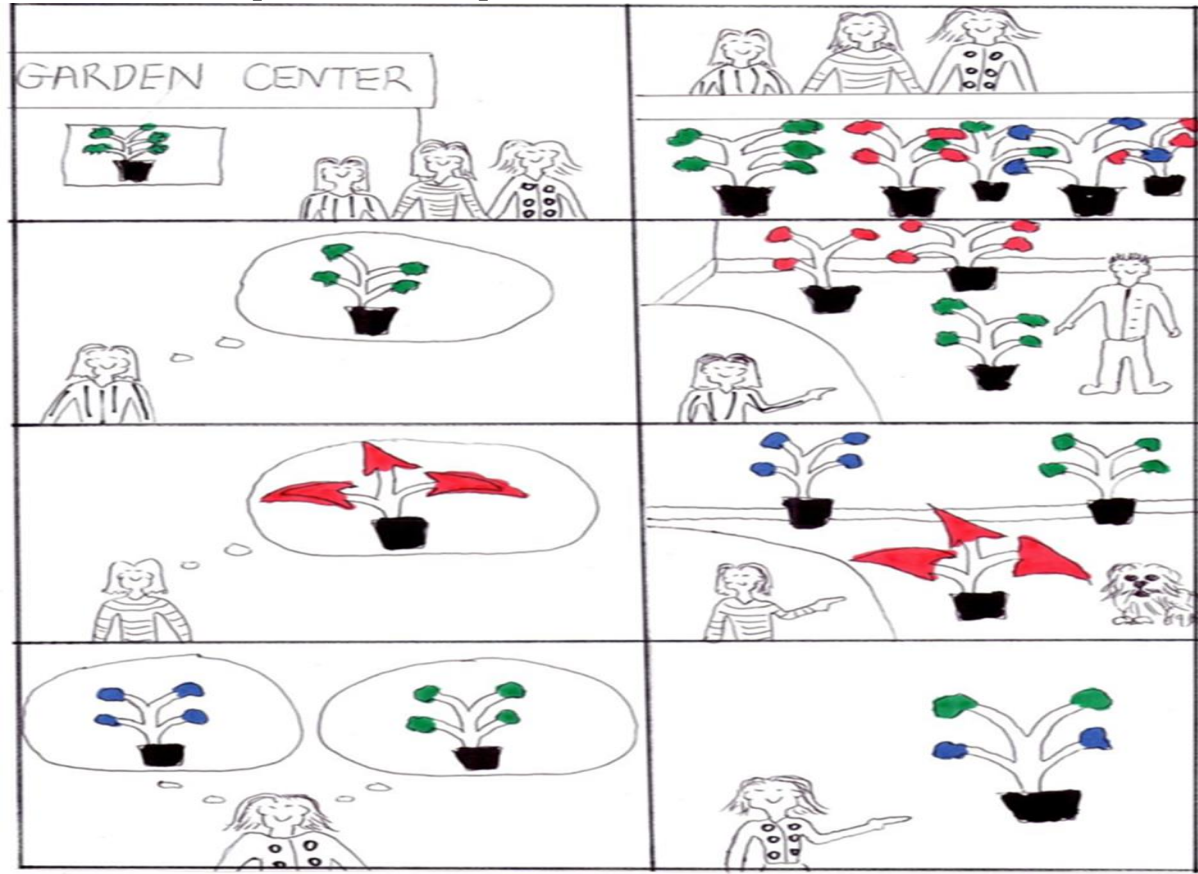
Narrative tasks used in the main study for the pre, post, and delayed tests:

1. **Pre-test (writing test, adopted from Hsu, 2015)**
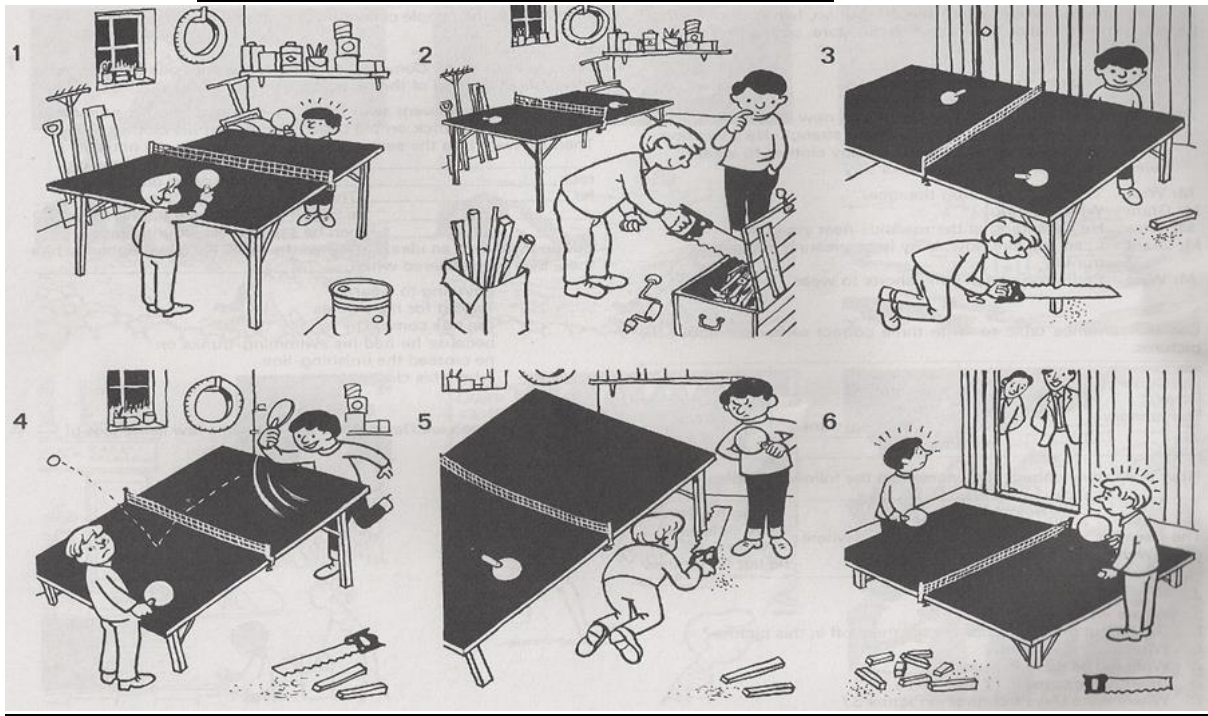
2. **Pre-test (speaking test, adopted from Hsu, 2015)**
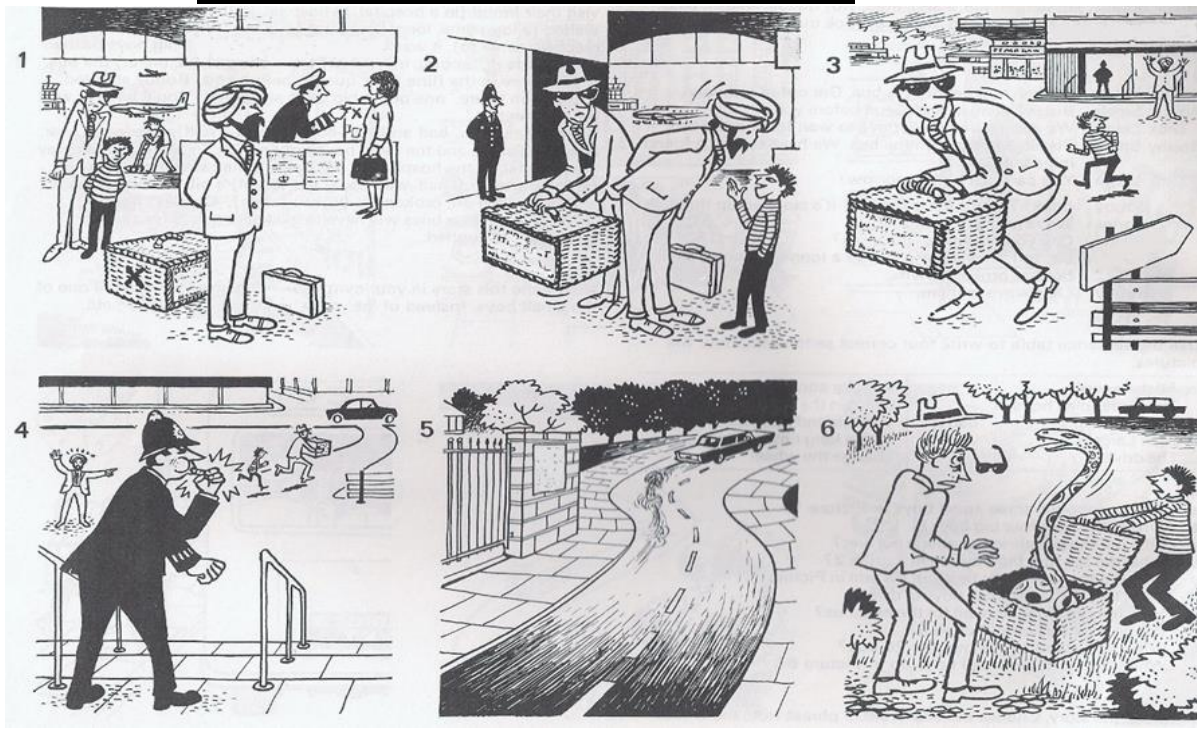
3. Post-test1 (writing test) (https://sites.google.com/site/winokure/picture-cuedstory-telling

4. Post-test1 (speaking test, adopted from Hsu, 2015)
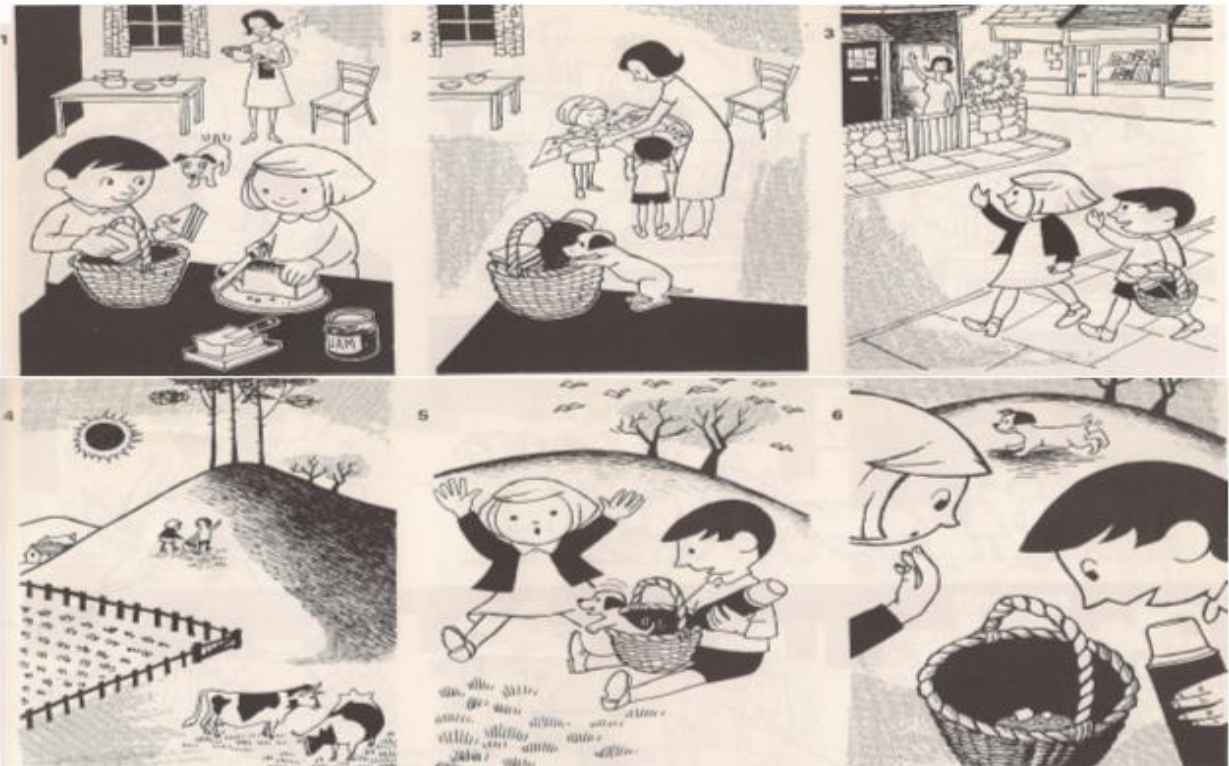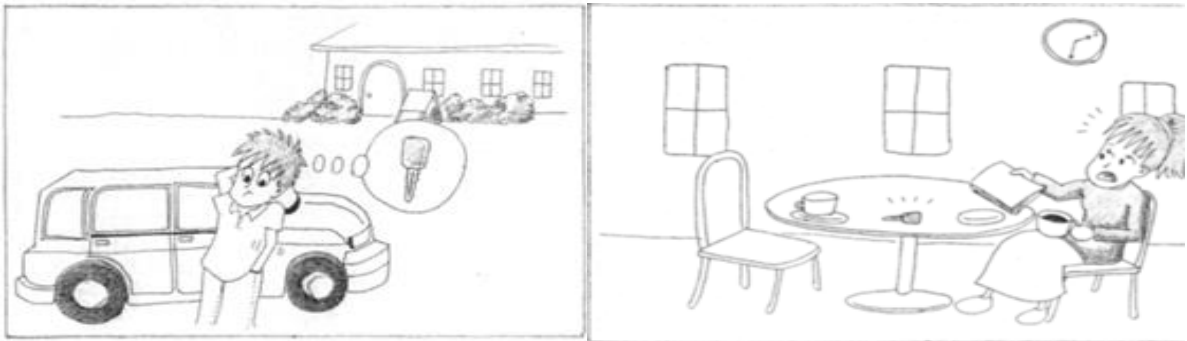
6. Delayed test 1 (speaking test, adopted from Hsu, 2015)
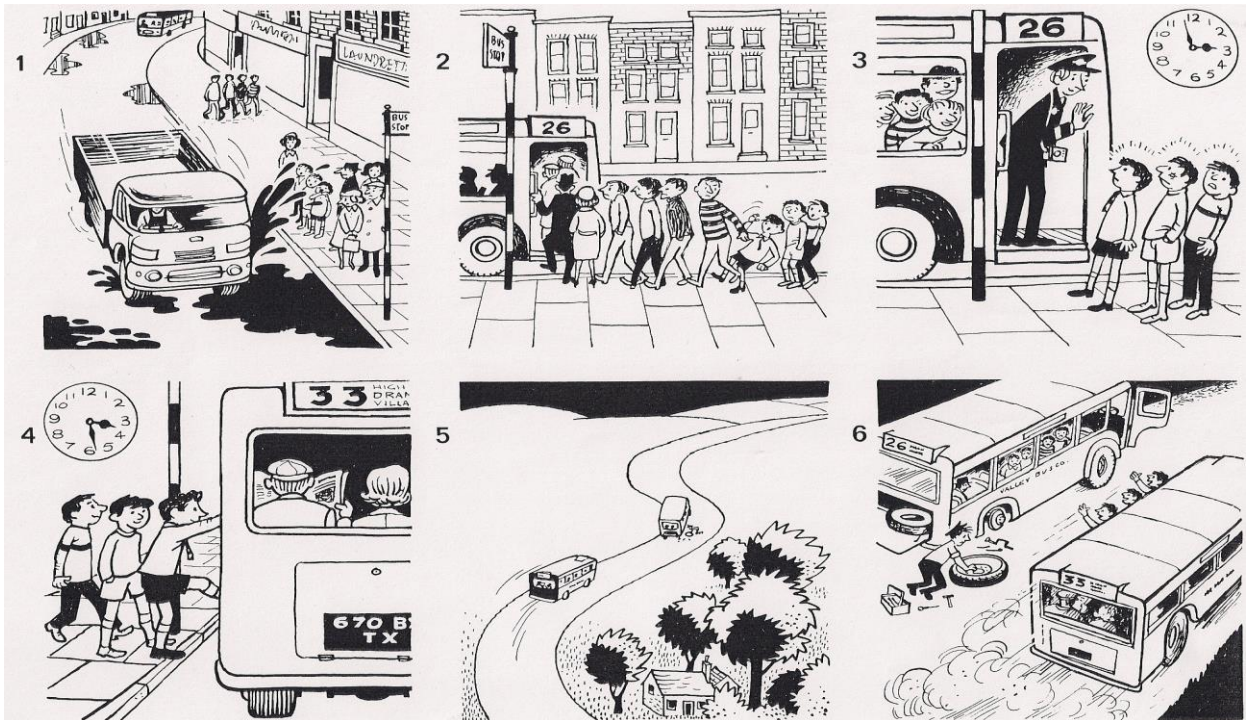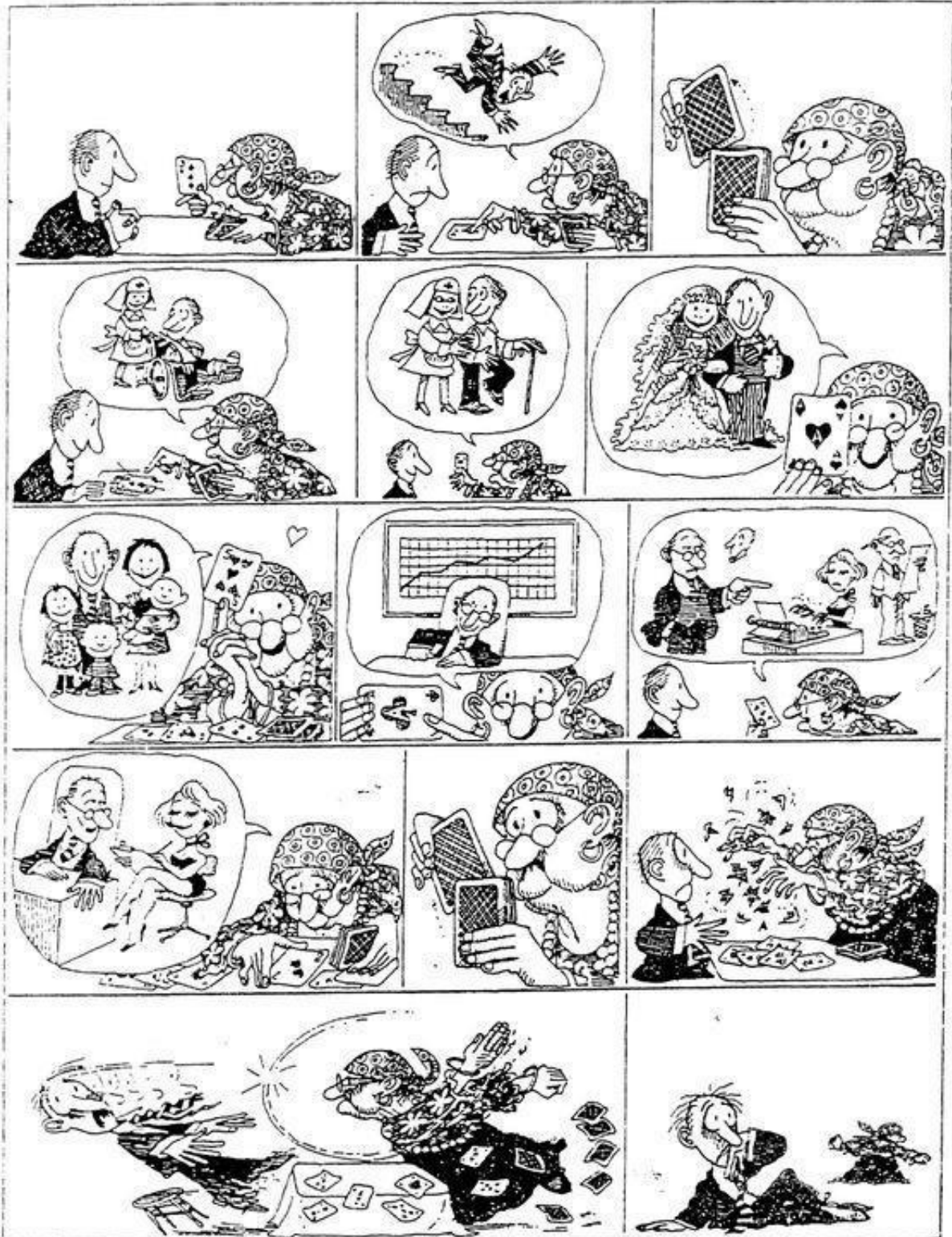
QUINO, ¡Qué mala es la gente!, 1996.

8. Post-test 2 (speaking test)

Ediciones de la Flor (@Ed_delaFlor) / Twitter
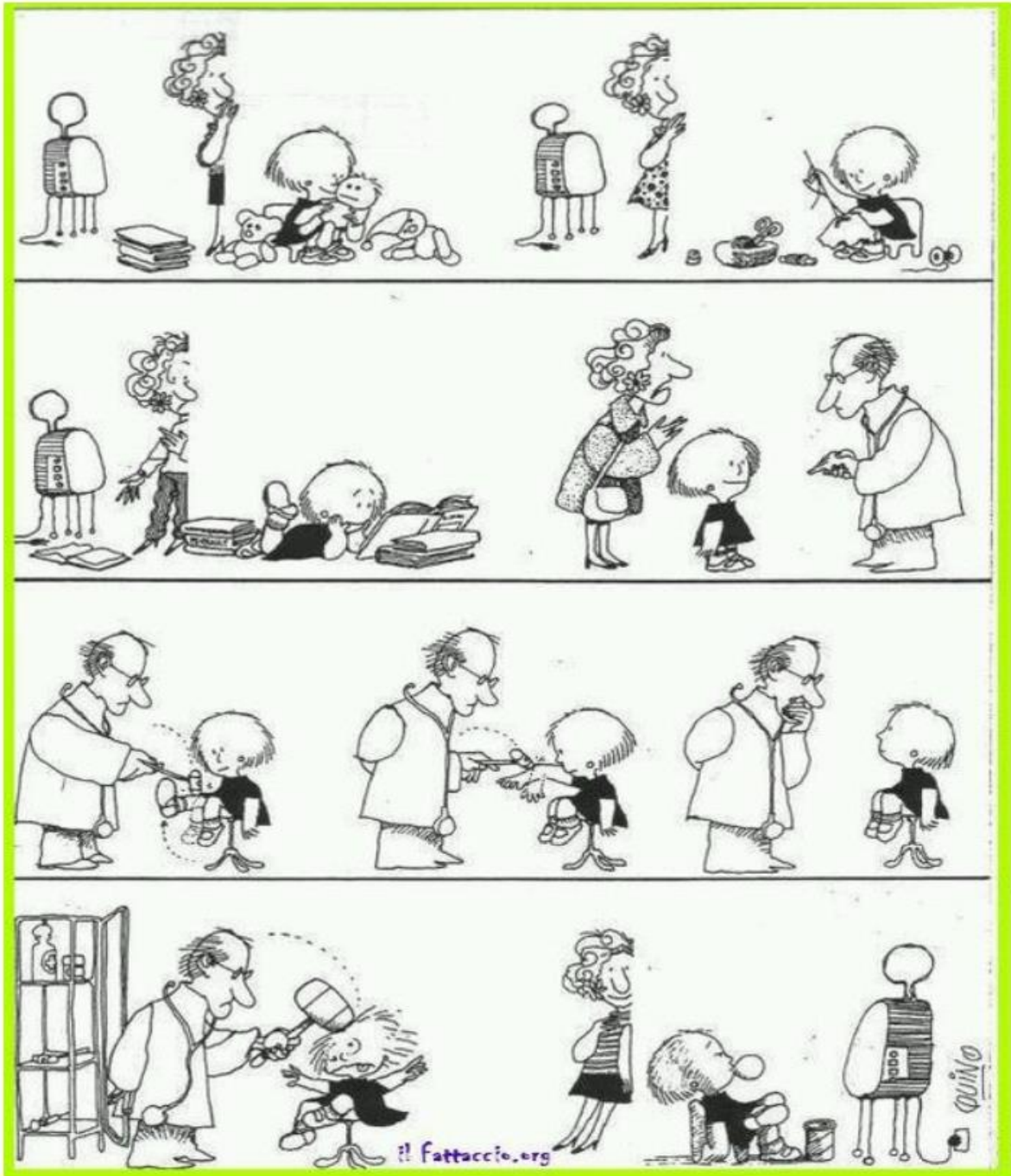
III.   <u>Instructions (for all the narrative tasks used in the tests):</u>

You will be given a set of pictures and you have only (**<u>10 minutes/1 minute</u>**) to write/tell a coherent and complete story about these pictures. If there is anything you are unclear about the task, please feel free to ask me before you begin.

After receiving the pictures, you can start immediately writing your narration on a separate Word document/telling your story.

Thank you!

IV. **Narrative tasks used during the online interaction (main study)**

    a. <u>**Text-based chat (pictures were given in the wrong order and were divided between the SL and his/her NS/ST interlocutor)**</u>
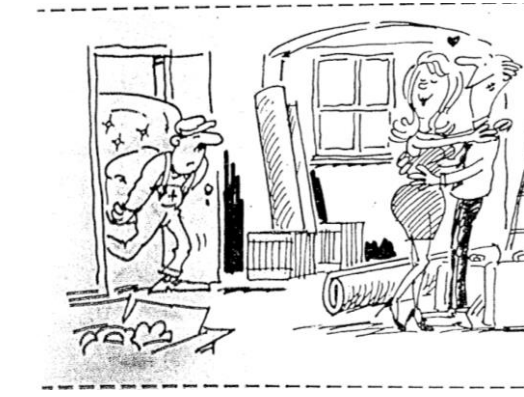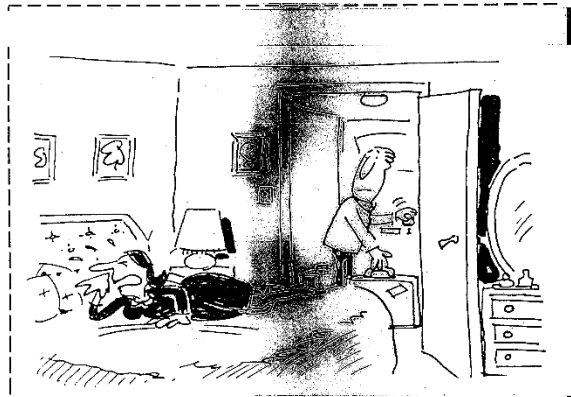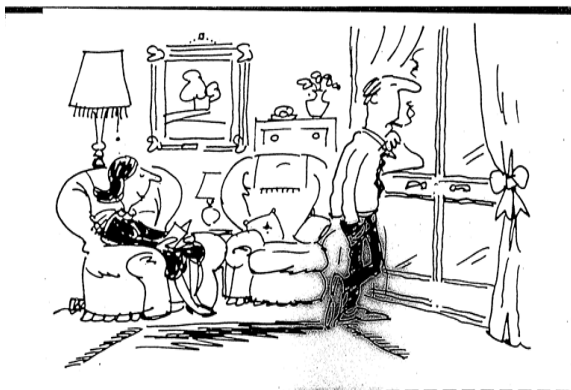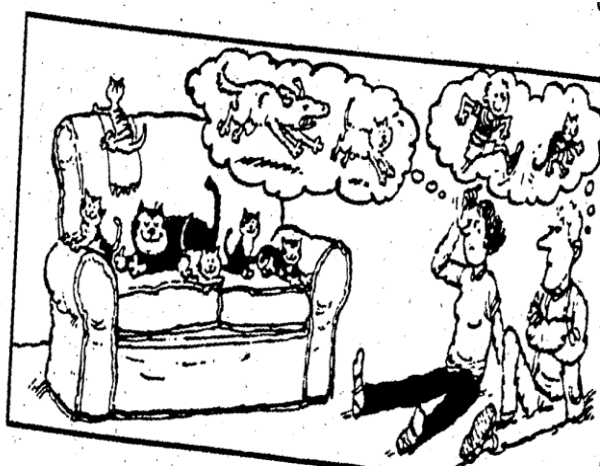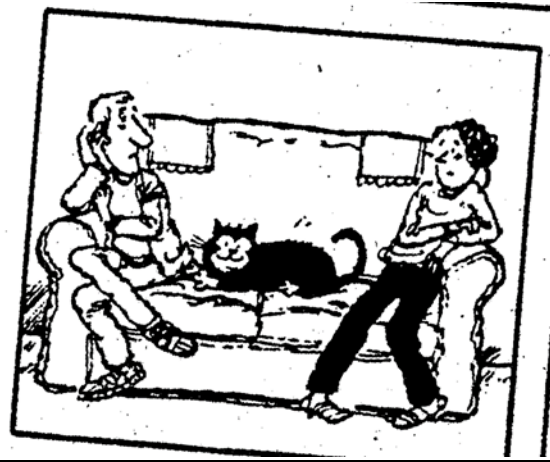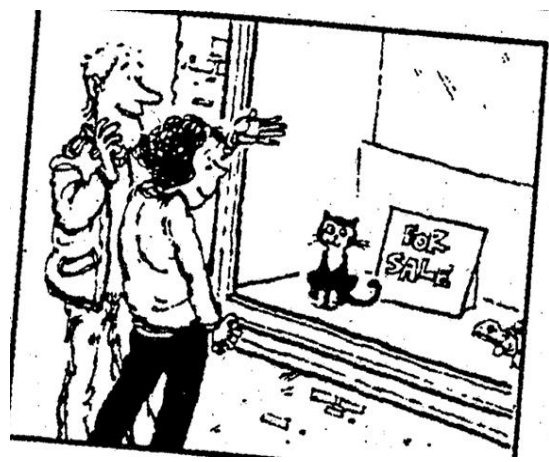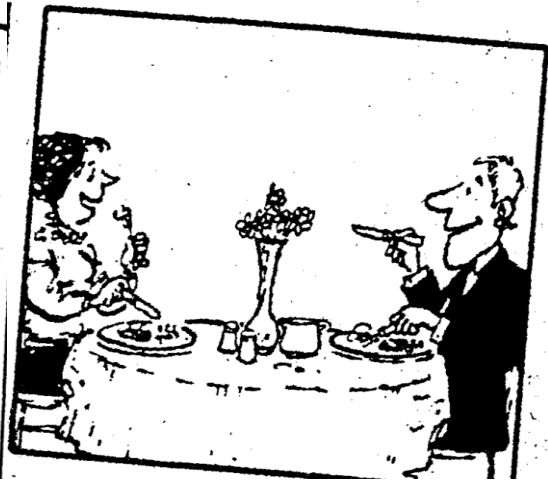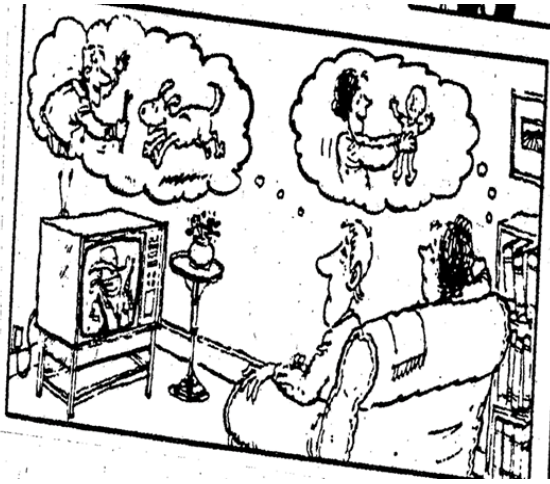
Instructions for guided pre-planning group (for SLs) (first draft).

1. Together with your partner try to describe each picture separately, so that you come up with one coherent story at the end.

2. You will have ten minutes to take notes and plan your task performance on a separate Word Document.

3. The Word Document should be closed before you start the session with your partner.

4. While planning;

   a. think of linking words and phrases such as firstly, besides, on the other hand, etc.

   b. think about grammar. For example:

      ➢ He likes the pair of shoes which looks trendy.

      ➢ She believes that her sister likes the dogs which have long ears.

Thank you

Instructions for guided pre-planning group (for SLs) (revised draft).

1. **1.** بالتعاون مع شريكك في الجلسة عليك ان تقوم بوصف مجموعة الصور (كلا على حدى) لتتمكنوا في النهاية من تشكيل قصة مترابطة المحتوى.

   Together with your partner take turns to describe each picture separately. Then you need to unjumble the pictures you have been given and put them in the right order so that you come up with one coherent story at the end.

2. **2.** لديكم 30 دقيقة كحد أقصى لانجاز المهمة الموكلة اليكم.

   You only have 30 minutes maximum to achieve the task goal.

3. **3.** قبل ان تبدأ الجلسة سيكون لديك 10 دقائق لتأخذ ملاحظات على ملف وورد وتخطط الطريقة التي ستعمل بها على شرح الصور وانجاز المهمة الموكلة لك.

   Before the session starts, you will have ten minutes to take notes and plan your task performance on a separate Word Document.

4. **4.** جميع الملاحظات يجب انو تكون مكتوبة باللغة الانكليزية.

   All your notes should be written in English.

5. **5.** لا داعي ان تكتب كل شيء بالتفصيل عندما تقوم بأخذ ملاحظات.

   Don't write everything in detail.

6. **6.** يجب ان تغلق ملف الوورد قبل أن تبدأ المحادثة مع زميلك في المحادثة.

   The Word Document should be closed before you start the session with your partner.

7. **7.** عند اخذ الملاحظات يجب ان تراعي الآتي:
   فكر بالكلمات والعبارات المفيدة والمرتبطة بمحتوى المهمة والتي يمكن لك ان تستخدمها لتجعل القصة أكثر متعة.
   فكر ايضا باستخدام عبارات لتنتقل من فكرة لأخرى مثل:

   First, second, etc.

فكر بالقواعد, تحديدا استخدام صيغ مثل:

> He like**s the pair of shoes <u>which looks trendy.</u>**
> She believe**s** that her sister like**s** the dogs **<u>which have long ears</u>**.

While planning
    a. think of varied useful expressions and words you could use to make your story look more interesting.
    b. think of linking words and phrases such as firstly, besides, on the other hand, etc.
    c. think about grammar; *in particular*, employing forms like
> He like**s the pair of shoes <u>which looks trendy.</u>**
> She believe**s** that her sister like**s** the dogs **<u>which have long ears</u>**.

8. لاتتردد بطلب المساعدة من زميلك اذا احتجت لذلك خلال الجلسة

During the task performance, don't hesitate to ask your partner for help if you need it.

9. اذا كان لديك اي سؤال او استفسارفيما يتعلق بالمهمة نفسها او الارشادات المرفقة قبل البدء بالجلسة من فضلك أسألني مباشرة.

If there is anything you are unclear about regarding the instructions or the task itself, please feel free to ask me before the session starts.

<div align="center">

شكرا لك
Thank you

</div>

Instructions for no pre-planning group (SLs working in the control group) (first draft).

1. Together with your partner try to describe each picture separately, so that you come up with one coherent story at the end.

Thank you

Instructions for no pre-planning group (for SLs) (revised draft).

1. بالتعاون مع شريكك في الجلسة عليك ان تقوم بوصف مجموعة الصور (كلا على حدى) لتتمكنوا في النهاية من تشكيل قصة مترابطة المحتوى.

   Together with your partner take turns to describe each picture separately. Then you need to put the jumbled pictures you have been given and put them in the right order so that you come up with one coherent story at the end.

2. لديكم 30 دقيقة كحد أقصى لانجاز المهمة الموكلة اليكم.

   You only have 30 minutes maximum to achieve the task goal.

3. لا تتردد بطلب المساعدة من زميلك اذا احتجت لذلك خلال الجلسة.

   During the task performance, don't hesitate to ask your partner for help if you need it.

4. اذا كان لديك اي سؤال او استفسار فيما يتعلق بالمهمة نفسها او الارشادات المرفقة قبل البدء بالجلسة من فضلك أسألني مباشرة.

   If there is anything you are unclear about regarding the instructions or the task itself, please feel free to ask me before the session starts.

شكرا لك

Thank you

Instructions for NSs & STs (first draft).

Together with your partner try to describe each picture separately, so that you come up with

one coherent story at the end.

Thank you

Instructions for NSs & STs (revised draft).

1. Together with your partner take turns to describe each picture separately. Then you need to unjumble the pictures you have been given and put them in the right order so that you come up with one coherent story at the end.

2. You only have 30 minutes maximum to achieve the task goal.

3. While working on the tasks, please try to help your partner by drawing his/her attention to the mistakes/errors they make with their English that you notice, especially those related to articles, tense, and use of relative clauses (For guided pre-planning group only).

4. Provide feedback wherever it seemed appropriate and in whatever form seemed appropriate during the interaction.
5. If there is anything you are unclear about regarding the instructions or the task itself, please feel free to ask me before the session starts.

Thank you

V. **Spot the difference tasks used during the piloting and the online sessions of the main study:**

   a. **Text-based chat**
   (https://svetlanaurisman.files.wordpress.com/2014/03/passive-pic.jpg)
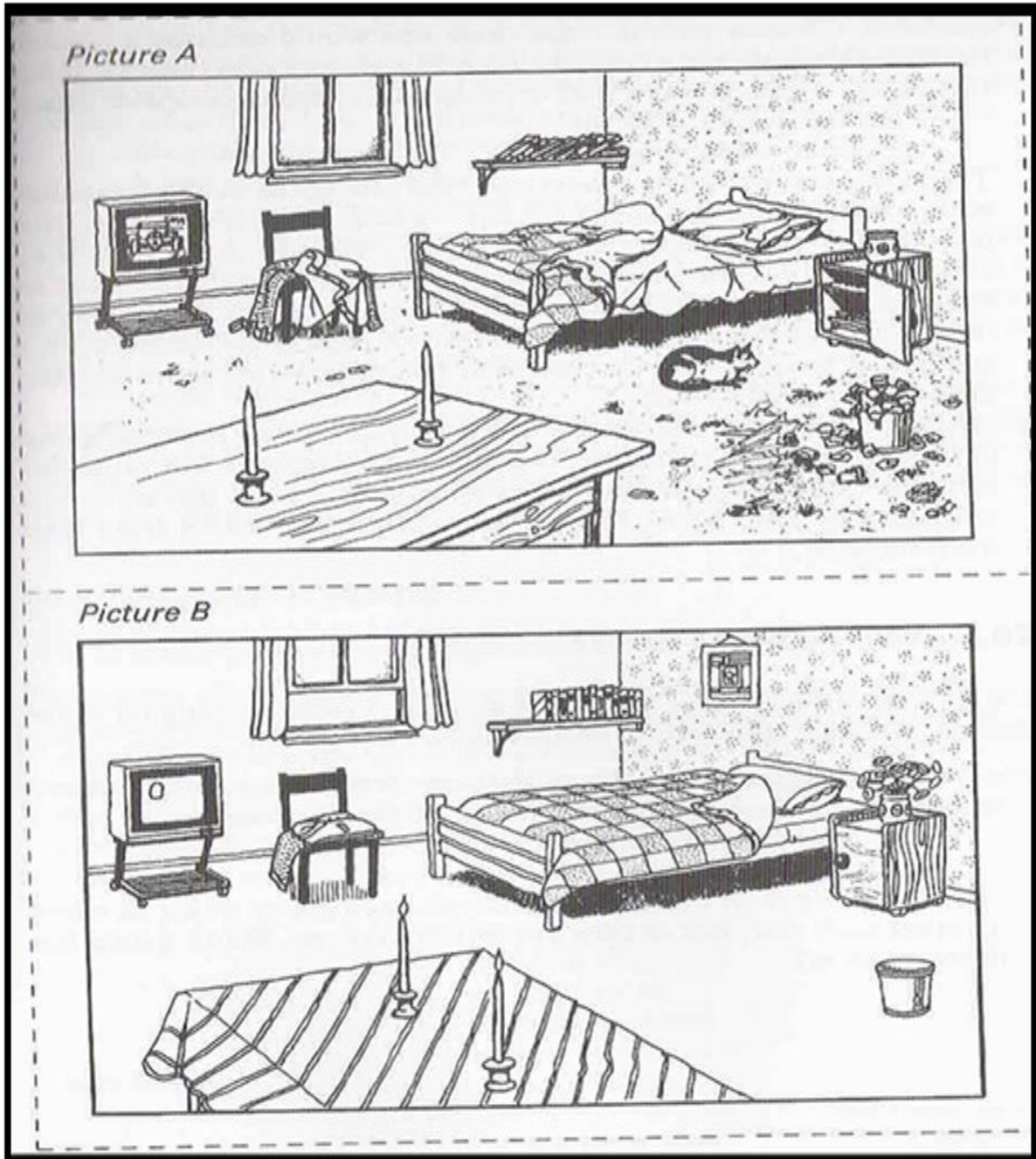
b. <u>**Voice chat**</u>
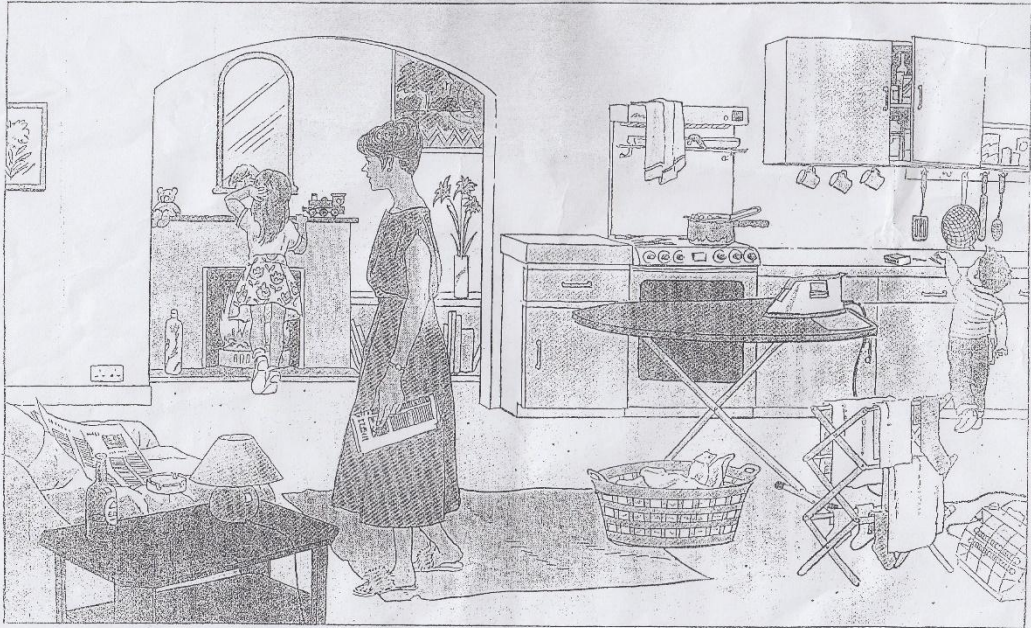
Student A

Student B
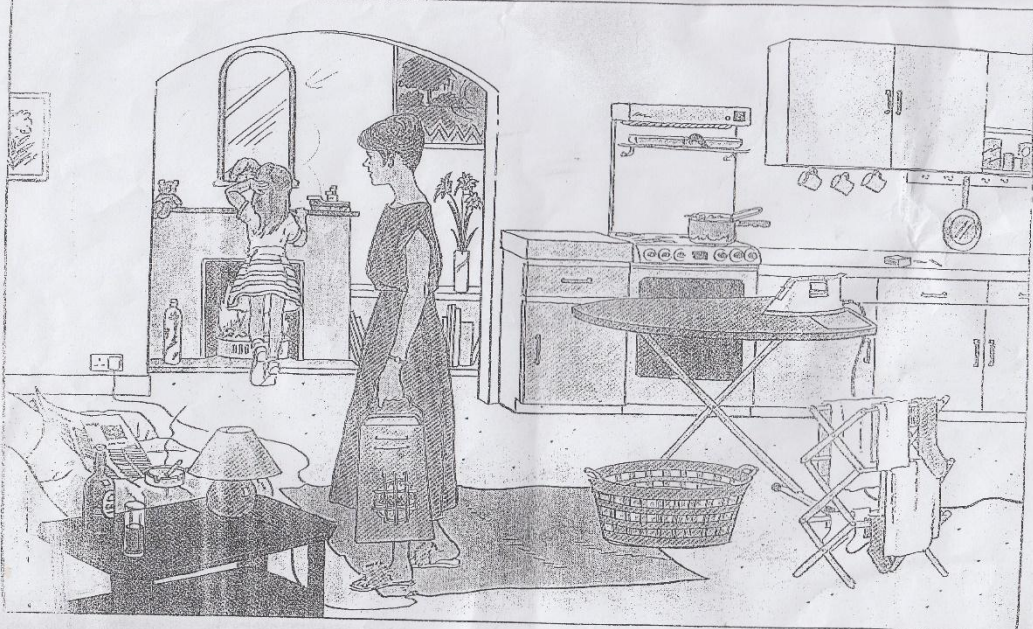
Instructions for guided pre-planning group (for SLs) (first draft).

1. Together with your partner try to describe the picture you each have to identify the differences between the two copies.
2. There is a minimum of 5 and a maximum of 12 differences.
3. You will have ten minutes to take notes and plan your task performance on a separate Word Document.
4. The Word Document should be closed before you start the session with your partner.
5. While planning;
   a. think of appropriate prepositions of place such on the right, to the left, etc.
   b. think about grammar. For example:
      ➢ In the middle, there is a basket which we use to put the laundry.
      ➢ The girl who is wearing a skirt and a shirt is standing in front of the mirror.

Thank you

Instructions for guided pre-planning group (for SLs) (revised draft).

1. عليك ان تعمل بالمشاركة مع زميلك على وصف الصورة التي لدى كل منكما وذلك لاكتشاف الفوارق الموجودة بين النسختين.

   Together with your partner try to describe the picture you each have to identify the differences between the two copies.

2. عليك ان تقوم بإعطاء معلومات مفصلة قدر الامكان للحصول على نتائج افضل والتمكن من تحديد كافة الفوارق عند وصف شيء ما لزميلك.

   For better interaction, try to provide as much detail as you could when describing an item to your partner.

3. يوجد 5 فوارق كحد أدنى و12 فارق كحد أقصى بين الصورتين.

   There is a minimum of 5 and a maximum of 12 differences.

4. لديك فقط 30 دقيقة لايجاد الفوارق بين الصورتين مع زميلك.

   You only have 30 minutes maximum to achieve the task goal.

5. قبل ان تبدأ الجلسة سيكون لديك 10 دقائق لتأخذ ملاحظات على ملف وورد وتخطط الطريقة التي ستعمل بها على انجاز المهمة الموكلة لك.

   Before the session starts, you will have ten minutes to take notes and plan your task performance on a separate Word Document.

6. جميع الملاحظات يجب انو تكون مكتوبة باللغة الانكليزية.

   All your notes should be written in English.

7. لا داعي ان تكتب كل شيء بالتفصيل عندما تقوم بأخذ ملاحظات.

   Don't write everything in detail.

8. يجب ان تغلق ملف الوورد قبل أن تبدأ المحادثة مع زميلك في المحادثة.

   The Word Document should be closed before you start the session with your partner.

9. عند اخذ الملاحظات يجب ان تراعي الـآتي:

فكر بالكلمات والعبارات المفيدة والمرتبطة بمحتوى المهمة والتي يمكن لك ان تستخدمها خلال الجلسة لوصف الاشياء وخصوصا تلك التي لاتعلم الكلمة المناسبة لها في اللغة الانجليزية.

فكر ايضا باستخدام عبارات لتنتقل من فكرة لأخرى مثل:

On the right, to the left, etc.

فكر بالقواعد, تحديدا استخدام صيغ مثل:

➢ In the middle, there is **a** basket **which we use to put the laundry**.
➢ The girl **who is wearing a** skirt and **a** shirt is standing in front of the mirror.

While planning;
   a. think of varied useful expressions and words you could use to describe an item to your partner, especially those you are not sure what they are called in English.
   b. think of appropriate prepositions of place such on the right, to the left, etc.
   c. think about grammar; *in particular*, employing forms like

➢ In the middle, there is **a** basket **which we use to put the laundry**.
➢ The girl **who is wearing a** skirt and **a** shirt is standing in front of the mirror.

10. لاتتردد بطلب المساعدة من زميلك اذا اجتجت لذلك خلال الجلسة
During the task performance, don't hesitate to ask your partner for help if you need it.

11. اذا كان لديك اي سؤال او استفسارفيما يتعلق بالمهمة نفسها او الارشادات المرفقة قبل البدء بالجلسة من فضلك أسألني مباشرة.
If there is anything you are unclear about regarding the instructions or the task itself, please feel free to ask me before the session starts.

شكرا لك
Thank you

Instructions for no pre-planning group (for SLs) (revised draft).

1. عليك ان تعمل بالمشاركة مع زميلك على وصف الصورة التي لدى كل منكما وذلك لاكتشاف الفوارق الموجودة بين النسختين.
Together with your partner try to describe the picture you each have to identify the differences between the two copies.

2. عليك ان تقوم بإعطاء معلومات مفصلة قدر الامكان للحصول على نتائج افضل والتمكن من تحديد كافة الفوارق عند وصف شيء ما لزميلك.
For better interaction, try to provide as much detail as you could when describing an item to your partner.

3. يوجد 5 فوارق كحد أدنى و12 فارق كحد أقصى بين الصورتين.
There is a minimum of 5 and a maximum of 12 differences.

4. لديك فقط 30 دقيقة لايجاد الفوارق بين الصورتين مع زميلك.
You only have 30 minutes maximum to achieve the task goal.

5. لاتتردد بطلب المساعدة من زميلك اذا اجتجت لذلك خلال الجلسة
During the task performance, don't hesitate to ask your partner for help if you need it.

6. اذا كان لديك اي سؤال او استفسارفيما يتعلق بالمهمة نفسها او الارشادات المرفقة قبل البدء بالجلسة من فضلك أسألني مباشرة.

If there is anything you are unclear about regarding the instructions or the task itself, please feel free to ask me before the session starts.

شكرا لك

Thank you

اذا كان لديك اي سؤال او استفسارفيما يتعلق بالمهمة نفسها او الارشادات المرفقة قبل البدء بالجلسة من فضلك أسألني مباشرة.

Instructions for NSs & STs (first draft).

1. Together with your partner try to describe the picture you each have to identify the differences between the two copies.

2. There is a minimum of 5 and a maximum of 12 differences.

Thank you

Instructions for NSs & STs (revised draft).

1. Together with your partner try to describe the picture you each have to identify the differences between the two copies.

2. For better interaction, try to encourage your partner to provide as much detail as they could when describing an item.

3. There is a minimum of 5 and a maximum of 12 differences.

4. You only have 30 minutes maximum to achieve the task goal.

5. While working on the tasks, please try to help your partner by drawing his/her attention to the mistakes/errors they make with their English that you notice, especially those related to articles, tense, and use of relative clauses (For guided pre-planning group only).

6. Provide feedback wherever it seemed appropriate and in whatever form seemed appropriate during the interaction.

.

7. If there is anything you are unclear about regarding the instructions or the task itself, please feel free to ask me before the session starts.

Thank you!

Instructions for guided pre-planning group (for SLs) (first draft).

1. The picture you have is a drawing of Richard's student room at university. Your partner also has a drawing of the same student room but his/her drawing is not complete. Help his/her to complete the drawing by saying where the things go. Your partner can ask you questions but must not see your drawing.
2. You will have ten minutes to take notes and plan your task performance on a separate Word Document.
3. The Word Document should be closed before you start the session with your partner.
4. While planning;
   a. think of appropriate prepositions of place such on the right, to the left, etc.
   b. think about grammar. For example:
      ➢ On the table, there is a small lamp which we use for study.
      ➢ .There are shelves where we put books just behind the door.

Thank you!

Instructions for guided pre-planning group (for SLs) (revised draft).

1. لديك وزميلك صورة لغرفة ريتشارد في الجامعة ولكن الصورة التي مع زميلك غير مكتملة. والمطلوب هو مساعدة زميلك على اكمال الصورة بإعطائه معلومات عن اماكن الاشياء كما تراها في صورتك. يحق لزميلك طرح الأسئلة لكن لايجب ان يرى الصورة ابدا.

   The picture you have is a drawing of Richard's student room at university. Your partner also has a drawing of the same student room but his/her drawing is not complete. Help his/her to complete the drawing by saying where the things go. Your partner can ask you questions but must not see your drawing.

2. للحصول على نتائج افضل, عليك ان تقوم بإعطاء معلومات مفصلة قدر الامكان عند وصف الاشياء وتحديد أمكانها لزميلك.

   For better interaction, try to provide as much detail as you could when describing an item and where it is in the room to your partner.

3. لديك فقط 30 دقيقة لايجاد الفوارق بين الصورتين مع زميلك.

   You only have 30 minutes maximum to achieve the task goal.

4. قبل ان تبدأ الجلسة سيكون لديك 10 دقائق لتأخذ ملاحظات على ملف وورد وتخطط الطريقة التي ستعمل بها على انجاز المهمة الموكلة لك.

   Before the session starts, you will have ten minutes to take notes and plan your task performance on a separate Word Document.

5. جميع الملاحظات يجب انو تكون مكتوبة باللغة الانكليزية.

   All your notes should be written in English.

6. لا داعي ان تكتب كل شيء بالتفصيل عندما تقوم بأخذ ملاحظات.

   Don't write everything in detail.

7. يجب ان تغلق ملف الوورد قبل أن تبدأ المحادثة مع زميلك في المحادثة.

   The Word Document should be closed before you start the session with your partner.

8. عند اخذ الملاحظات يجب ان تراعي الـآتي:
   فكر بالكلمات والعبارات المفيدة والمرتبطة بمحتوى المهمة والتي يمكن لك ان تستخدمها خلال الجلسة لوصف الاشياء وخصوصا تلك التي لاتعلم الكلمة المناسبة لها في اللغة الانجليزية.

فكر ايضا باستخدام عبارات لتنتقل من فكرة لأخرى مثل:

On the right, to the left, etc.

فكر بالقواعد, تحديدا استخدام صيغ مثل:

- On the table, there **is a** small lamp **which we use for study**.
- There **are** shelve**s** **where we put books** just behind the door.

While planning;

a. think of varied useful expressions and words you could use to describe an item to your partner, especially those you are not sure what they are called in English.
b. think of appropriate prepositions of place such on the right, to the left, etc.
c. think about grammar; *in particular*, employing forms like

- On the table, there **is a** small lamp **which we use for study**.
- .There **are** shelve**s** **where we put books** just behind the door.

9. لاتتردد بطلب المساعدة من زميلك اذا احتجت لذلك خلال الجلسة

During the task performance, don't hesitate to ask your partner for help if you need it.

10. اذا كان لديك اي سؤال او استفسارفيما يتعلق بالمهمة نفسها او الارشادات المرفقة قبل البدء بالجلسة من فضلك أسألني مباشرة.

If there is anything you are unclear about regarding the instructions or the task itself, please feel free to ask me before the session starts.

Thank you!

Instructions for no pre-planning group (for SLs) (first draft).

The picture you have is a drawing of Richard's student room at university. Your partner also has a drawing of the same student room but his/her drawing is not complete. Help his/her to complete the drawing by saying where the things go. Your partner can ask you questions but must not see your drawing.

Thank you!

Instructions for no pre-planning group (for SLs) (revised draft).

1. لديك وزميلك صورة لغرفة ريتشارد في الجامعة ولكن الصورة التي مع زميلك غير مكتملة. والمطلوب هو مساعدة زميلك على اكمال الصورة بإعطائه معلومات عن اماكن الاشياء كما تراها في صورتك. يحق لزميلك طرح الأسئلة لكن لايجب ان يرى الصورة ابدا.

The picture you have is a drawing of Richard's student room at university. Your partner also has a drawing of the same student room but his/her drawing is not complete. Help his/her to complete the drawing by saying where the things go. Your partner can ask you questions but must not see your drawing.

2. للحصول على نتائج افضل, عليك ان تقوم بإعطاء معلومات مفصلة قدر الامكان عند وصف الاشياء وتحديد أمكانها لزميلك.

For better interaction, try to provide as much detail as you could when describing an item and where it is in the room to your partner.

3. لديك فقط 30 دقيقة لايجاد الفوارق بين الصورتين مع زميلك.

You only have 30 minutes maximum to achieve the task goal.

4. لاتتردد بطلب المساعدة من زميلك اذا احتجت لذلك خلال الجلسة

During the task performance, don't hesitate to ask your partner for help if you need it.

5. اذا كان لديك اي سؤال او استفسارفيما يتعلق بالمهمة نفسها او الارشادات المرفقة قبل البدء بالجلسة من فضلك أسألني مباشرة.

If there is anything you are unclear about regarding the instructions or the task itself, please feel free to ask me before the session starts.

شكرا لك

Thank you!

<u>Instructions for NSs & STs (first draft).</u>

> The picture you have is a drawing of Richard's student room at university. Your partner also has a drawing of the same student room but his/her drawing is complete. Ask your partner questions to help you complete the drawing by saying where the things go.

Thank you!

<u>Instructions for NSs & STs (revised draft).</u>

1. The picture you have is a drawing of Richard's student room at university. Your partner also has a drawing of the same student room but his/her drawing is complete. Ask your partner questions to help you complete the drawing by saying where the things go.

2. For better interaction, encourage your partner to provide as much detail as they could when describing an item.

3. You only have 30 minutes maximum to achieve the task goal.

4. While working on the tasks, please try to help your partner by drawing his/her attention to the mistakes/errors they make with their English that you notice, especially those related to articles, tense, and use of relative clauses (For guided pre-planning group only).

5. Provide feedback wherever it seemed appropriate and in whatever form seemed appropriate during the interaction.

6. If there is anything you are unclear about regarding the instructions or the task itself, please feel free to ask me before the session starts.

Thank you!

VII.   Decision-making tasks used in the pilot and main study

a.   Text-based chat

Task 1 (Decision Making Scenarios Worksheet | Decision Making | Applied Psychology (scribd.com): Lind has just returned to school and has been out of the study habit for 7 years. She has found it very difficult to get back into the habit of studying. Her time is further stretched by responsibilities of being a wife and a mother of two pre-school aged children. Based on the scenario, together with your partner **identify the problem(s)**. **Discuss all possible solutions** as well as **the pros and cons (if any) for each solution**.

Task2 (https://www.michigan.gov/documents/mdch/Decision_Making_Activities_456247_7. doc?cv=1&session-id=40f4fffec79348f799cd56fe530db050):

Six patients need a heart transplant. Below you are given information based on the circumstances described for each person. There is only one heart donor at this time. All of the patients are eligible to receive this heart. All are physically able. And all have compatible tissue and blood typing. Which patient would you choose to receive the heart? Why?

You and your partner must agree on the choice.

It is worth mentioning that patients who do not receive this heart will not automatically die. Some (not all) will probably survive until another donor is available.

Patient Waiting List

- ☐   31 year old male; brain surgeon at the height of his career; no children

- ☐   12 year old female; accomplished violinist; blind

- ☐   40 year old male; teacher, 2 children

- ☐   15 year old female; unmarried, 6 months pregnant

- ☐   17 year old female; waitress; high school dropout; supports/cares for a brother who is

       severely disabled.

- ☐   38 year old female; AIDS researcher; no children; lesbian

Discussion Prompts given to NSs/STS **ONLY**:

1)   What was your decision?

2)   How did you arrive at your decision?

3)   How was your decision influenced by your values? Attitudes? Prejudices?

Task 3

You and your partner have a very important and big project that is due tomorrow and you haven't even started it yet. You plan on spending several hours doing it tonight. However, your friends suddenly ask you to go to a concert tonight. They won tickets to your favorite group and want you to go with them. You need to get the project done, but you really want to go to the concert with your friend.

Together with your partner you need **to define the problem**, **explore the alternatives**, **consider the consequences**, **and identify your values**. Based on your discussion, *__you should come up with your final decision__*.

b. **Voice chat**

Task 1

Your friends, (including someone you have a crush on) start smoking. What would you do?

Together with your partner you need **to discuss the reasons that you think might make them take that decision**, **consider the consequences**, **and identify your values**. Based on your discussion, *__you should also think of ways to convince your friends to give up on that bad habit.__*

Task 2: You've been offered the chance to move abroad to a country where you could build a better future. However, if you take this chance, you will be leaving your whole life behind. What would you do?

Together with your partner you need to **discuss the pros and cons of living in another country; consider the short-term/long-term effects, if any, this decision might have on your personal/social life.** Based on your discussion, *__you should take the decision of whether to stay or leave.__*

Task 3

You've just started working for a software start-up. One day you have a great idea for a new app, which you believe is going to revolutionize the industry. You share it with James, your coworker. The following day, you find out James had presented the idea to the boss, claiming it was his. Needless to say, the boss loved it, gave James the credit and the promotion, and the app is going to be produced soon.

Together with your partner discuss the reasons that you think might make James do what he did, think of ways/solutions to convince the boss that the idea was yours. What would you do if all your attempts to convince the boss were in vain? Are you to blame for what happened? You might also discuss any learned lessons?

Your friends are bullying someone at university/work. You used to be friends with the person that is being bullied. What would you do?

Together with your partner discuss the following:

> **What bullying means and what kind of actions does it involve?**
> **Why do you think some people like to bully others?**
> **Any short term/long term effects you think this action could have on the person who's being bullied?**
> **Would you try to convince your friends to stop that action? How?**
> **What would you do if they refused to listen to you?**
> **How your personal values and beliefs affect your decision?**

a. **Planning guidelines for the planning group:**

The aim of note-taking is to help you organise your ideas, thoughts, etc., before the online session starts. It is important that you are only making notes. ***Remember***, they are only there to help you speak; you are not supposed to read a speech that you've just written.

In order to make the most of the time given to you prior to task performance, there are certain stages you need to consider:

الهدف الاساسي من اعطائك الوقت لاخذ ملاحظات قبل الجلسة هو مساعدتك على تنظيم أفكارك. من المهم جدا ان تكون على شكل ملاحظات مقتضبة وحسب. تذكر ان الغاية من الملاحظات هي مساعدتك لتتحدث وليست مجرد كلام كتبته وستقرؤه للتو.

لمساعدتك على استغلال ذلك الوقت الذي ستأخذه قبل البدء بأي جلسة والاستفادة منه يتوجب عليك تتبع الخطوات التالية:

1. يجب ان تتأكد من ما هو مطلوب منك خلال الجلسة (وصف, رواية قصة, حل مشكله, الخ).

   Be clear on the task goal (describing, narrating, problem-solving, etc.)

2. أن تأخذ ملاحظات لا يعني ان تضيع وقتك بكتابة جمل كاملة وصحيحة لأنك غالبا ستنساها فيما بعد . انما هي عمليه فقط لكتابة كلمات اساسيه قد تساعدك على تذكر والتحدث عن افكار معينة خلال الجلسة مع زميلك.

   Taking notes does not mean writing full sentences, rather it's only jotting down keywords that will spur certain thoughts at the moment because as you start interacting with your partner, it's highly likely that you will not recall all the sentences you've written earlier.

3. لاتضيع وقتك بأخذ ملاحظات بكل الكلمات التي قد تراها في الصورة لأنك ستحتفظ بالصورة طوال فترة الجلسة.

   Don't take notes of all the words you can see in the picture. This could waste your time as you'll still have access to the task materials during the session.

4. خصص بعضا من الوقت للتفكير بقواعد اللغة التي يمكن استخدامها كالزمن على سبيل المثال (ماضي, حاضر, مستقبل)

   Think about the most appropriate tense to use to achieve the task goal (past, present, future).

5. خصص بعض الوقت ايضا للتفكير بمفردات او عبارات قد تفيدك عندما تتحدث مع زميلك لتبدو أكثر طلاقة في اللغة (عبارات الانتقال من جملة او فكرة لأخرى).

   Think of certain vocabulary/ phrases/ useful expressions that could make your writing/speech look more fluent (e.g., transition signals)

6. عندما تشعر انك غير قادر على ايجاد الكلمة المناسبة باللغة الانكليزية لشيء ما, حاول ان تفكر بمرادفات او عبارات قد تساعدك على شرح ذلك الشيء لزميلك حتى يتمكن من مساعدتك واعطائك الكلمة التي تبحث عنها.

When you feel that you can't find/don't know the right word in English for a particular item, try to think of synonyms/phrases (e.g., what it is used for?) that could help you explain that item to your partner, so that s/he could be able to tell you the correct answer e.g., container, deep, put food in, cook food= saucepan

b. **Example of notetaking:**

Describe a trip you took last year. You should say where you went, what you did, and why you travelled there. What made the trip memorable?

Note taking: Egypt: → Cairo (a week) (food, architecture)

Al-Gardaka (3 days) (nature, nightlife)

Why: renew visa (primary)

    visit friends, holiday (secondary)

Memorable, wife's/husband first time; New Year's celebration

c. **Screenshot 1: An example of learner's notes during the planning time:**

bullyin is abuse either physical or emotional

rough childhood - abusive parents - wrong parenthood-ing - mentally ill and doesn't realise it , they bully other people so they can see them feeling down and they would feel better

it depends on how he/she handle it , cause depression , physical harm maybe

yes , ask for the reasons first , what do u get out of it , is this how you want people to treat you/your kids

stand up for them if they refuse , bully them let them have a taste of their own medicine , maybe file a complaining paper to the boss/school for inapproprate behaviour   www.thundershare.net

affects it in a decent amount cuz YOLO and people don't deserve to live happy and not worry if they'll get bullied when they get out and live their lives

I. **Interview questions for the stimulated recall interviews with the SLs:**

1. ‏كيف وجدت الارشادات؟
   How did you find the instructions?

2. ‏كيف تقيم الوقت الذي أعطي لك قبل ان تبدأ الجلسة
   ‏مفيد جدا
   ‏مفيد
   ‏لست متأكدا
   ‏غير مفيد
   ‏غير مفيد اطلاقا
   ‏ولماذا...

   Did you find the time given to you to take notes before the online session starts helpful?
   How helpful or unhelpful the planning time was?   **(Revised)**
   A. Very helpful
   B. Helpful
   C. Neither helpful nor unhelpful/not sure
   D. Unhelpful
   E. Very unhelpful
      And Why……………………………………………………………………?

3. ‏اي من الاشياء المذكور ادناه حاولت تغطيته او التركيز عليه خلال الوقت الذي اخدته قبل الجلسة؟
   ‏القواعد
   ‏المعنى
   ‏اختيار الكلمات
   ‏التهجئة
   ‏اللفظ
   ‏..............................شيء اخر (الرجاء التحديد) ولماذا
   Which things on the list below you covered during the planning time?

   A. Grammar
   B. Meaning
   C. Vocabulary/word choice
   D. Spelling
   E. Pronunciation
   F. Other, please specify, and why?………………………………………….

4. ‏كيف وجدت المهمة التي اوكلت اليك اليوم؟
   How did you find the task?

5. اي من الاشياء المذكور ادناه حاولت تغطيته او التركيز عليه خلال الجلسة؟

القواعد
المعنى
اختيار الكلمات
التهجئة
اللفظ

....شيء اخر (الرجاء التحديد) ولماذا

Did you try to focus on grammar as well? **(Original)**

Which things on the list below you covered during the online session? **(Revised)**

  G. Grammar
  H. Meaning
  I. Vocabulary/word choice
  J. Spelling
  K. Pronunciation
  L. Other, please specify……………………………………………………………

6. من فضلك اخبرني كيف كنت تشعر عندما كنت تتحدث مع زميلك خلال الجلسة.

How chatting with a person you've never met made you feel? **(Original)**

Please tell me how you felt interacting with the person you were chatting with. **(Revised)**

7. كيف تشعر وانت تعلم انك تتحدث مع شخص يتحدث الانجليزية كلغته الام/يدرس اللغة الانجليزية؟

How did you feel about communicating with NS/ST?

8. كيف تشعر وانت تعمل على المهمة الموكله لك بطريقة الكتابة/المكالمة الصوتية؟

How did you feel about performing the task in the written/oral mode?

9. ماذا برأيك افضل صفات التواصل باللغة الانجليزية, ان وجدت, باستخدام التكنولوجيا الحديثة باستخدام الكتابة/ او المكالمة الصوتية؟

What do you think are the best features if any, about interacting via text chat/audio chat?

10. ماذا برأيك اسوء صفات التواصل باللغة الانجليزية, ان وجدت, باستخدام التكنولوجيا الحديثة باستخدام الكتابة/ او المكالمة الصوتية؟

What do you think are the worst features, if any, about interacting via text chat/audio chat?[24]

---

[24] It is worth noting that these general questions were asked in addition to other specific ones according to each participant's interaction. For instance, when a specific incidence of error correction was identified, the participant was asked the following questions: "When you were doing the task, I noticed your partner said.. (an example of correction)... could you tell me about that? What were you thinking about at this instance? What was going in your mind when you typed "X"? Then I asked follow up questions depending on what s/he said, such as how helpful/unhelpful s/he felt it was? Whether s/he understood or not? etc.

II. <u>**Interview questions for the semi-structured interviews with NSs/STs**</u>

1. Please tell me how you find the online experience.

2. Please tell me how you felt interacting with your partner via text-based/voice chat.

3. Please tell me how you felt interacting with the person you were chatting with.

4. How well did you feel you worked with your partner?

5. Please tell me how you found the tasks. Why?

6. How effective do you think the feedback you gave was? Why?

7. Why did you try to avoid explicit feedback? Could you tell me more about this, please?

8. Tell me the moments during the online exchanges you felt were particularly helpful or confused.

I. **Chat log samples:** A chat log excerpt for a typical voice chat session, taken from Majd (SL) and Rosy's (NS) interaction[25]:

Typist	Dialogue
Rosy:	Hello Majd, how are you?
Majd:	I'm fine, you?
Rosy:	I'm good, thanks
	So Majd what do you think bullying means?
Majd:	well, bullying is being eh maybe verbally or physically abusive, eh to eh other people?
Rosy:	ok?
Majd:	and maybe (pause), yeah it's like being harmful whether in a verbal way or in a physical way to eh other people
Rosy:	ok could you also describe that, erm tell me what kind of actions it involves more specifically?
Majd:	well, you can bully people eh maybe at school. There's also cyber bullying nowadays eh on social media apps and stuff like that. And also maybe like eh in our task we have (pause) someone is bullying someone at university or work so it could be (pause) a physical abuse maybe eh at school or, or (pause) verbal abuse.
Rosy:	okay, erm so actions that are hurtful to someone, or actions that harm someone physically you mean or harm them emotionally?
Majd:	well, it can be either eh emotionally or physically.
	Yeah exactly
Rosy:	erm so like calling someone names?

---

[25] Pseudonyms were used to further make learners' identities.

Majd: yeah calling someone names or maybe making fun of
                 someone; the way they dress or maybe make fun of eh the
                 way (pause) their finger shaped or maybe (pause) the
                 lifestyle that they live, eh it can be a lot of actions.

Rosy: okay and do you think that anything is different with the
                 way someone might bully people at university versus in
                 the workplace?

Majd: (pause) erm, maybe not really eh but eeh (pause) it would
                 be (pause) much more different eh in a workplace like it
                 can be, erm you can either make fun of eh the way they
                 work or eh if someone is very eh eh maybe eh work
                 isolated eh (pause) they may, they might make eh
                 someone might make fun of them?

Rosy: what do you mean by if they're work isolated?

Majd: I mean eh like someone eh would like to maybe he is not
                 a very eh team player. He likes to work on his own.
                 maybe?

Rosy: ok just a quick language point, you wouldn't say he is
                 very team-player, you would say he is; if someone is not
                 really a team-player or not very big on team-playing,
                 acting as a team. Erm yeah I see what you mean, so
                 maybe you could also bully someone by isolating them?

Majd: yeah, sure eh you mean like making your own group and
                 like rejecting or like keeping someone out of it?

Rosy: yeah, erm so erm why do you think that some people
                 like to bully others?

Majd: well, eh there is I think there is a lot of (pause) reasons
                 maybe one of them is eeh (pause) like eh (pause) maybe
                 they were eh like a rough childhood? they had a rough

	childhood? Maybe abusive parents, yeah like they had abusive parents so, so when they go to maybe like school or university or something eh they like to take eeh that anger out on other kids and slash people?
Rosy:	ok, I see what you mean. So maybe they're, it's a person who has experienced something quite erm, quite traumatic in their life?
Majd:	yeah
Rosy:	erm, so just thinking is there a way to phrase it using more like a relative clause?
Majd:	(pause) erm, sorry I don't I didn't get what you said last
Rosy:	I'm just trying to erm focus more on the language aspect, so could you rephrase an of those sentences using relative clauses? like the kind of person who likes to bully others might be a person who…?
Majd:	okay, so the kind of person that would eh bully other people eeh could be a person who, who is a per, is a person who had a rough childhood or an abusive parents or (pause) maybe
Rosy:	yeah
Majd:	okay
Rosy:	I see what you mean. Erm it could be maybe people who are kind of inherently insecure with themselves.
Majd:	yeah
Rosy:	so people who feel better about themselves when they put other people down?
Majd:	yeah exactly. They like to see other people down so they feel better.

Rosy: yeah, I think I agree with you. Erm so speaking about

the person who is being bullied, erm so in this scenario, it's someone you used to be friend with. Erm so obviously what kind of long-term and short-term effects do you think erm they'll experience from being bullied?

Majd: well, I think it could eh, eh ca. Well it depends on the

way they like eh handle it? So some people just ignore and eh like can deal with the hate or deal with the bullying, some people stand up for themself, but if eeh we're talking about someone who is not standing up for themself then it could maybe cause eh physical harm if it was a physical abuse, maybe it could cause eeeh (pause) a depression? If like someone knew that they have to go everyday to work and just like somehow get abused by eh s, eh other group at work. So it could cause depression or like a mentally eh a mental illness.

Rosy: yeah, I see what you mean. Quick language notes you

would say it would cause depression not a depression. I think you said it correctly the second time but the first time you had the article. And also erm we wouldn't say if it's a physical abuse. We just say if it's physical abuse, so no article with abuse in this kind of a sentence, but when we want to talk about it in terms of like countable thing, we could say the actions, like an action erm but yeah. So erm yes I agree with you on that, yeah especially the physical harm it could get erm worse and worse and become more and more violent. So it could be a danger. Erm and yeah I think erm in terms of depression yeah it could definitely maybe, it could maybe have a negative impact on the person's self-esteem?

Majd: yeah sure

Rosy: erm, so there's an expression as well in English, so we

could say so if, for certain people with their personality maybe they're able to sort of brush it off, or shake it off like they don't let it impact them and they don't let it kind of sink in. Erm, so like you said it just depends on how this person handles it. So are they the kind of person who is more sort of naturally resilient, or the

	kind of person who is quite sensitive to bullying actions, bullying behaviour.
Majd:	yeah
Rosy:	erm, so obviously the impact will probably be like negative, erm so if it's your friends who are doing the bullying, erm do you think you would try to intervene?
Majd:	yeah, of course like if eh especially eeh like the task say eeh they are bullying someone I used to be friend, friends with, so maybe I eh maybe I would first advice eh that friend, old friend of mine to maybe stand up for themself. If eh (pause) they are not eh standing up for themself maybe or maybe if, maybe I can eh, eh like talk to these group of people who are abusing him or her and stand up for them. Maybe I can eh or maybe I can just go out of my way and eh bully them, let them have a taste of their own medicine
Rosy:	ah, okay you would become a bullier, so that you can show them how it feels
Majd:	yeah, exactly
Rosy:	okay
Majd:	put them in their shoes
Rosy:	yeah, put them in their place or put them in his shoes
Majd:	yes, exactly
Rosy:	erm, I see what you mean. Well, it could kind off escalate things if it's a bit violent or if there's a lot of sort of back and forth but I see what you mean it might teach them a lesson erm.
Majd:	yeah, or maybe I can file like eeeh (pause) com eh, complaining eeh order or complaining paper to the maybe the boss at work or maybe at school?
Rosy:	yeah you could say file a complaint
Majd:	file a complaint yeah exactly

Rosy: so one more quick language point when you're talking about erm your friend, erm or former friend standing up for themselves instead of themself, you can say himself or herself but then with them, themselves

Majd: ah, ok

Rosy: okay, I think those were quite a few different ways of convincing your friends, so which do you think will be the most effective way to stop them to stop your friends, erm speaking to the person who's being bullied to fight back, or you're bullying your own friends, or speaking to them and trying to persuade them to stop.

Majd: well, maybe I will try to eh advice my old friends to eh stand up for themselves first, then maybe like if not eh if he or she eh wouldn't do such a thing maybe I could eh advice, them to stop bullying them, or talk to those group of people about stop eh about stopping the bullying?

Rosy: so, but it's kind of difficult situation that you'd be in because you're putting your relationship with your friends at risk for a person who is not your friend currently.

Majd: well, I think it's worth it if they are that type of a friend, (pause) or that type of a person like who would like to bully eeh other people and get a laugh out of it just like I eh that's not a good person to be around in general so…

Rosy: so you'd sort of, maybe that would connive you to not be a friend with them, you think?

Majd: yeah, maybe if they wouldn't stop doing eh such a thing

Rosy: okay, and so what would you do if they refuse to listen to you?

Majd:	well. Like I said I might eh like eh (pause) file a complaint eeh and eh like to the school (pause) or maybe the boss.
Rosy:	okay, so just to take it to the next stage in terms of the power, higher your power chain
Majd:	yeah, exactly since like they are not listening to me and they keep on doing so, so..
Rosy:	okay, so how would you describe your personal values and beliefs in making this decision?
Majd:	well, I believe that everyone deserves to be happy and eh do their everyday routine without eh having to worry about like "I am going to school now, now I have to put up with everyone bullying me saying that and that about me, this and that about me eh", or maybe like just eh going to work because you only live once and everyone deserves to be happy with their life so and not like I said be worry about eh other people bullying him or maybe ruining their day.
Rosy:	ok, so would you say, would you describe yourself as a person who has a strong sense of right and wrong?
Majd:	yeah, I would
Rosy:	and do you feel you have, you are a kind of a person who has a strong sense of loyalty to friends
Majd:	yeah, even if I am not eeh a friend with that person anymore eeh I still would do it for the old time sake.
Rosy:	okay so for the principle of the situation
Majd:	yeah
Rosy:	okay, erm do you think the bully randomly chooses their victim?
Majd:	well, maybe they would go for, like I said, like the weak one or maybe they would go for the isolated type of person eh, eh

347

or maybe they could also like go for the one who has maybe eh deformity or a disability just like making fun of them or just bullying them in general

Rosy: ok, so like a person who is different for some reason from the norm?

Majd: yeah, exactly in their of course point of view he is different or she is different in their like point of view

Rosy: okay, when you choose to talk to the person who's being bullied, do you think they'd be able to defend themselves?

Majd: well, eeeh (pause) maybe like maybe they are weak to take a reaction, but maybe they need a push. Maybe they need eh (pause) eh someone to like eh get them going, maybe they need, someone like to hear, to hear it from someone or like someone to feel that it's not okay and you can stand up for yourself and not everyone is accepting what these other people are doing so (pause). I think it's not, eh If you don't get a reaction maybe like I said you would then talk to the bullies eeh (pause) group

Rosy: yeah, I somewhat agree with you. I think it's a tough one because yeah the person being bullied is feeling quite weak. Erm the may feel quite powerless but it might help them to feel that someone has their back or someone is encouraging them. Erm that someone believes in them, so maybe that would empower them and make them feel you know more courageous to stand up but it's hard when it's like a physical bullying, erm I think it's really hard to know whether it's better to suggest like in a way it could be deescalated but if you encourage them to like fight back or be violent in return it can make things erm worse perhaps so it's kinda like however you erm define standing up for themselves.

Majd: maybe if it's eh physical abuse or physical bullying maybe eh it's not, eh it's better not to go like and tell them to fight back or something, but maybe (pause) take it like you said to a

higher power or a higher eeeh people eh with higher power and just like maybe eh (pause) use eh these people and tell them what eh what you've been going through and maybe get help through these people

Rosy: yeah, do you think those victims might become bullies themselves as a kind of reaction to what they've been through in their life?

Majd: yeah, absolutely

Rosy: yeah, I think that sometimes that what happens especially when people are younger, sort of if people have been bullied themselves when they were pretty little erm then maybe later on they replicate that to feel powerful. It changes their personality slightly. Erm, Mohammed do you think there might be any positive impact on the personality of those victims in the long run?

Majd: well, it just like maybe eh get them eh some hint of what is coming up because like eh where , no matter where you go there's always a bully or you'll always be made fun of or something, so yeah it could maybe shape some stronger or maybe like eh you said make them more resilient and more eh maybe have the ability to shake it off more than usual.

Rosy: I think I agree with you that maybe it could yeah if they've been bullied and they come out of it alright, it could sort of teach them that you know they can get through anything, any bad experience because in life you'll always be faced with challenges and then you have to try to stay strong and believe in yourself. Okay I think that was it. Thanks Majd, it was a pleasure to work with you. Wish you all the best

Majd: same here. Thanks for your time and all the best for you too

Rosy: thanks, take care, bye

Majd: you too, bye

Chat log samples: A chat log excerpt for a typical text chat session, taken from Rima (SL) and Rana's (ST) interaction:

Typist	Dialogue
Rima:	Hi Rana
	How are you?
Rana:	Hi
	Fine
	How are you?
Rima:	Great
Rana:	Shall we begin?
Rima:	IS Everything OK?
Rana:	Yes, thanks for asking
Rima:	Sure
	Great
	Let's start
Rana:	Would you like to start describing your first photo?
Rima:	Sure
	Just a second
Rana:	Ok
Rima:	So the first photo
	Is in the living room
Rana:	Ok
	What is in there?
Rima:	There are two people
Rana:	Ok
	A man and a woman?
	How do they look like?

Rima: so I have a man and a weman

The man is sitting on a sofa

And I think he is watching Tv

Rana: Alright

Rima: The woman is standing on the right side

Rana: Mine has a man and a woman

Rima: Beside the door and she looks angry

Rana: On a date probably

Rima: Ok

Rana: Maybe in a café

And there is some dancing in the background

Rima: Good

Rana: The woman is elegant with a blonde hair

Rima: Great, what is she wearning

Rana: A dress

Rima: Good

Rana: A sleeveless short dress

Rima: My second Photo is like they bought a house

And there are some workers carrying furnitures

Rana: Why do you think they bought a house?

Rima: Carpet

Rana: Ah ok

Rima: And the room looks untidy and that leads they are (pause) putting

things for their new house

Rana: Ok

Who is in the picture?

Rima: The woman who has a blond hair

And the man

Rana: Ok

Rima: They are hugging each other

Rana:	Ok
	In my second photo, it seems like there is a party
Rima:	Whose there?
Rana:	A loud party with lots of dancing and drinking
	Maybe it is a club
Rima:	Good
Rana:	There is a woman with a black hair
Rima:	So, the man is dancing too?
Rana:	No
Rima:	What does she do?
Rana:	The man is not here
	The woman is alone wearing her long sleeves black dress
	And she is dancing happily
Rima:	Great
Rana:	With a different man, I guess
	Your turn
Rima:	Like boyfriend?
Rana:	I do not believe so
Rima:	My third photo is in a bedroom
	The woman with black hair is lying on the bed and crying
	And the man is getting ready to leave her, I guess
	He is carrying his bag
Rana:	Ok
	That's it?
Rima:	Yeah just that
Rana:	I have in my third photo the man walking in the street
	And the blonde woman is in front of him
Rima:	Are they walking
Rana:	It looks like he is following her
Rima:	Is she angry?

Rana: Like he wants to introduce himself to her

               No

Rima: Oh

               Yes, yes.

Rana: That's it

               Your turn

Rima: Does she look satisfied?

Rana: No, he has not reached her yet

Rima: I think I didn't write it correctly

               Ok

Rana: He is just behind her

Rima: Fine

               My turn

Rana: Ok

Rima: My fourth one looks like a party

               The man is dancing with the women

               Blondy women

Rana: Ok

Rima: And she is dancing too

               They look really happy

               Delighted

Rana: Ok

               My forth one is in the living room

               The woman with the black hair and the man

               The man on the right is staring into the window

               He looks depressed

Rima: Great

Rana: The woman on the other side is sitting on the armchair reading something

               OK

               Let's order them

Rima:	So now we sould order them
Rana:	Yes
Rima:	I think the woman with black hair is his ex-wife
	Or sth like that
Rana:	Yes
Rima:	So they didn't get on well
	With each other
Rana:	Yes, and he met the other woman
Rima:	Yeah
Rana:	which made his wife sad
	Then she managed to get over this terrible experience
	And went to a party
Rima:	So I think he decided to leave here
Rana:	Yeah
	But which photo is the first one?
Rima:	I think it is yours
	When he is staring
	Toward the window
Rana:	Yes, I think so too
	Then
Rima:	And then my first one when they look (pause) argured
	In the living room
Rana:	Alright
	Are they arguing?
	Do you think so?
Rima:	They just don't look and see to each other
	Yes, I think they had an argument
Rana:	Alright
	Then he walks down the street
Rima:	Yeah could be

Rana:	And sees that blonde woman
Rima:	Possible
Rana:	Then asks her for a date
	Then tells his wife he is leaving
Rima:	That true
	Then his wife get sad
	And went to the club
Rana:	Does the party in your photo looks like a wedding?
	Yeah
Rima:	No
Rana:	Ok
Rima:	I don't think so
Rana:	Maybe they went to the party on another date
	Then move in together to a new house
Rima:	It can be
Rana:	I think this is logical
Rima:	So the house is the last picture
	I completely agree
Rana:	Yeah
	Our time is over
Rima:	Right
Rana:	We will talk next week
	Take care
Rima:	Sure
	Take care
	Bye
Rana:	Bye

■■

1. **A Complete Interview Script with Majd:**

Researcher: How did you find the instructions?

Majd: كتير كانوا واضحين لان كانت الافكار اللي بدنا نحكي فيها مقسمة بشكل واضح ومفهوم

Researcher: How helpful or unhelpful the planning time was?
    F. Very helpful
    G. Helpful
    H. Neither helpful nor unhelpful/not sure
    I. Unhelpful
    J. Very unhelpful
     And Why…………………………………………………………?

Majd: كمان كان كتير مفيد وكنت فعلا بحاجته لاقدر نظم افكاري عن هالموضوع اللي كان بحاجة لكم جديد ومختلف من الكلمات اللي بتتناسب مع هالموضوع..بتخيل لولا وجود هالوقت قبل كل جلسة كانت مفرداتي وافكاري رح تكون كتير محدودة وبحس ادائي رح يكون اقل من ناحيه اللغة المستخدمة (رح تكون ابسط الموضوع اكيد ماكان دايما سهل وسلس واكيد كان عم يتأثر بنوع المهمة وموضوعها والخلفية اللي عندي عن كل موضوع. اخد النوتس لما يكون في صور اسهل لان في شي مجسد قدامك وبدك تحكي عنه ماكتير بحاجة تفكر وتطلع افكار اكيد لما بيكون ماعندي خلفية مسبقة او افكار عن الموضوع بتكون مسألة انه لاقي كلمات وافكار تناسب الموضوع اصعب بكتير لهيك بفضل بوقتها او بنجبر تكون جملي قصيرة ومقتضبة

Researcher: Which things on the list below you covered during the planning time?
    M. Grammar
    N. Meaning
    O. Vocabulary/word choice
    P. Spelling
    Q. Pronunciation
    R. Other, please specify, and why?…………………………………………..

Majd: الوقت استخدمته بشكل أساسي لنظم افكاري عن موضوع التنمر اللي متل ماسبق وقلت انو بحاجة ل مخزون كبير من التعابير والمصطلحات لتقدري تتناقشيه

Researcher: How did you find the task?

Majd: الموضوع من ارض الواقع ومتداول بشكل كبير هالفترة سواء بحياتنا اليوم او على الانترنت

Researcher: Which things on the list below you covered during the online session?
- S. Grammar
- T. Meaning
- U. Vocabulary/word choice
- V. Spelling
- W. Pronunciation
- X. Other, please specify…………………………………………………………

Majd: اكيد طبيعة المحادثة الصوتيه مابتعطيك هالاريحية لتفكري باللي بدك تقوليه قبل ماتقولي وانما بيطلع الكلام عفوي اكتر وهاد اللي بيفسر الوقفات احيانا والتلبك لما تضيع الفكرة او المفردة المناسبة بس مع هيك طبعا كنت حريص انه صحح اي خطأ بكتشفه بلغتي وانا عم احكي اما بإعادته مره تانية او بجملة جديدة بتحتوي على الاستخدام الصح

كمان صرت انتبه عالملاحظات اللي عم تعطيني ياها شريكتي لما كنت ركب الجملة بشكل غير دقيق وكانت تعلمني كيف ممكن قول الجملة متل ما البريطانين بيقولوها

Researcher: Please tell me how you felt interacting with the person you were chatting with.

Majd: كنت كتير مرتاح وعم احكي ع طبيعتي اكيد هاد الشعور تولد بالتدريج مو من اول جلسة

Researcher: How did you feel about communicating with a NS?

Majd: أكيد هالشي خلاني حاول انتبه اكتر ع لغتي سواء من ناحيه القواعد لاختيار الكلمات لتركيب الجملة بشكل صحيح بحيث تكون موصله المعنى المطلوب

Researcher: How did you feel about the feedback that was given to you by your partner?

Majd: انا بعرف انه لغتي فيها اخطاء اكيد وكنت مستعد لخوض هالتجربة بكلشي فيها مع هيك صراحة مابخفيكي انه الاحساس بالبداية ماكان حلو لما كانت تصلحلي كنت حس انه مستوايي اضعف مما تخيلت يمكن لان ماكنت بعرفها منيح بالبداية وكنا عم نحكي كتابه مع هاد كله مابخفيكي انه كتير استفدت من تعليقاتها

Researcher: How did you feel about performing the task in the written mode?

Majd: هي حلوة يعني متل ماحكيت الواحد بيقوي حاله بالنسبة للقواعد والكلمات وتركيب الجمل بشكل صحيح اكيد التواصل مع شريكتي بطريقة الكتابة كان بمثابة وسيلة لكسر الحواجز بيني وبينها قبل ماننتقل للمحادثة الصوتيه. هلق بظن كمان صارت تعرف اكتر عن مستوايي اللغوي وانا صرت مدرك اكتر وين بتكمن نقاط ضعفي

Researcher: How did you feel about performing the task in the oral mode?

Majd: كان في بعض من التوتر اكيد لفكرة انه شو ممكن تكون نوعيه التاسك بهل الحالة.. هل رح اقدر ادي التاسك بشكل صحيح ولاقي كلام مناسب قوله لمدة نص ساعة او لا..مع بداية الجلسة وسماع صوت شريكتي مشاعر التوتر والقلق بدأت تخف شوي شوي

الحقيقة كمان كان في توتر كبير خلال الجلسة لما تضيع مني الكلمات بالرغم انها تكون حرفيا ع راس لساني

Researcher: What do you think are the best features if any, about interacting via text chat?

Majd: كنت عم اقدر اخد بالملاحظات اللي تعطيني ياهم (بما انها صلحتلي اكتر من مرة) بما يتعلق باستخدامي ل اللغة لان المعلومات قدامي عالسكرين وبقدر برجعلها بأي وقت هالشي خلى المعلومة تترسخ اكتر براسي لاستخدمها بشكل صحيح بعدين بتخيل هالشي صعب يتوفر لو كنا عم نحكي صوت بس

كمان التهجئة تبع الكلمات لان عم شوفهم عالسكرين ..اكيد كمان فكرة لفكر انه معي وقت بالجملة وركبها بشكل صحيح او اقرب للصحيح قبل مابعتها

Researcher: What do you think are the best features if any, about interacting via voice chat?

Majd: انك تسمعى اللفظ الصحيح للكلمات بس بنفس الوقت الكلام بيكون عفوي اكتر يعني ماعندي وقت كتير لفكر واكتب وامسح الكلام بيطلع بشكل لحظي
انا حبيت الكتابة بس بالنسبة الي بفضل التحدث اكتر لحتى اتدرب عالكلام اكتر ويكون عندي طلاقة باللغة اكتر.. كانت فرصتي انه احكي قد مابقدر خاصة انه كنت عم اتواصل مع شخص اجنبي..بشكل عام الكتابة ممكن تقعدي عالكمبيوتر وتمارسيها مع اي حدا بس التحدث مع شخص اجنبي هاد الشي اللي نوعا ما كان خارج عن المألوف

Researcher: What do you think are the worst features, if any, about interacting via text chat/audio chat?

Majd: كمان اللي لاحظته هو تقبلي الشخصي لاسلوبها بالتصحيح بحاله الكتابه يمكن كنت حس اني intimidated

بس لما سمعتها عم تصححلي بالمحادثة الصوتيه كان الموضوع جدا طبيعي وعادي يمكن هي وحدة من مساوئ المحادثة الكتابيه انك مابتحس كتير بالطريقة اللي عم ينحكى فيها الكلام والمقصد منه

Researcher: How did you find the whole experience?

Majd: انا كتير كنت متحمس للفكرة ومهتم انه طور لغتي كان الشي السلبي الوحيد هو توقيت الجلسات احيانا يكون غير مناسب الي. غير هيك انا كتير حبيت الموضوع وكتير استفدت سواء قواعديا او من ناحية المصطلحات والمفردات. اول مادخلت عالتجربة حاولت سقف التوقعات تبعي يكون منخفض لحتى ما كتير اتفاجأ ويخيب املي بس بصراحة التجربة فاقت سقف توقعاتي بكتير

.. اللي كان بميز هالجلسات هو التعليقات اللي كنت اخدها من شريكتي سواء ع لغتي او ع اختيار الكلمات يعني صرت لما استخدم هالكلمات كون على ثقة انها صحيحة. بالبداية كنت خاف احكي شي غلط لحتى مابين ضعفي اللغوي قدامها بس بعدين راح هالخوف وصرت احكي حتى لو مو متأكد من صحة اللي عم احكيه لتكون الفائدة اكبر

****

2. A Complete Interview Script with Rosy

Researcher: How did you find the online experience?

Rosy: Yeah, I found it very interesting. Overall, I felt the experience was positive. I enjoy taking part and some of the tasks were quite funny and quite interesting to hear the students' perspectives and their ideas. However, It's difficult to do the tasks when we can't type, I felt. I was struggling a bit with my ability to concentrate on the learners erm sort of what does the topic that we are talking about and getting the task done at the same time and trying to address language issues. So, the first part where we could type, I felt I was able to keep track a bit better because I could see visually what they were writing, erm when it's just auditory I found it a little bit more difficult.

Researcher: How did you find interacting with the learners via text chat and voice chat?

Rosy: Well, I did like being able to hear their voices and hear sort of how they formulate things in a spoken mode, but when you are not in front of them like able to write things down in front of them or even to text while speaking. I do find it hard to sort of interrupt them. For example, during one of the sessions, I interrupted the learner to correct him; like I did not want to wait too long but I sort of broke his concentration; he lost his train of thought for a bit, so it's always hard to find the balance between disrupting them and always drawing their attention to errors"

Researcher: Have you used any of these two modes for teaching purposes before? When and how?

Rosy: I have but kind of with mixed sort of forms a bit. So I've done sort of webinars with students where I give the students different tasks but it's all typing and I've done some skype chat with students. So yes, I mainly did work in the written mode, so I am not as used to the skyping with audio.

Researcher: What do you think about online teaching/learning?

Rosy: In the last few years, I've been quite interested in online learning and I've been studying it for the DELTA and I've been trying to develop more online work at my university where I work. So it's a part of the services that we offer; we are trying to do more online but definitely I think the one with the audio, kind of teaching or Skype teaching, I think it is challenging and after this experience I still fell a lot of challenges related to it, so still very positive I think a lot can be done in this way, but it's hard when it's spontaneous and when you are trying to accomplish a task. Like from a teacher's perspective, there was not a lot of sort of structure to the lessons, like I did not know how much I should try to get them to respond, so the task was left quite open in terms of how I should treat them as students so I wanted to try to be more implicit about the teaching as much as I could but I sometimes think maybe I should have done it in a different way to be more effective. So I think online teaching is tough when you are not entirely sure what your goals are or how the students want to learn online. So, I think if I was designing the course, I think this is something I would need to think through more for myself.

Researcher: How? Could you elaborate on that?

Rosy: With students and with online courses, they always say like with any other course you should do needs' analysis and you should find out what kind of learning style they like. I think maybe I was a bit reluctant because I did not really know the students very well, they have a good level of English, so was not sure how much sort of interrupt their flow. Since they were adults, I did not feel it is comfortable to kind of drilling them or getting them to like repeat back to me. So, there's a certain structure to it that it would be different if it was sort of a whole course with like a clear syllabus and kind of different aspects that it is not as a parrot when it's more sort of broken up into tasks like this.

Erm, I mean it is nice that it's open, but then it made me worry that I felt I was not sure what I was doing was as helpful or what is the most helpful thing to do. So maybe after the task having a chance to get feedback from students and say did you learn expressions, did you find this

helpful; were sort of more aware of your errors. I am sure you were getting this feedback from them but from a teacher's perspective I felt a little bit like I was sort of going through this blindly. I was not sure how much they were enjoying it, or how useful they were finding my suggestions and my ideas for their language. So I think having a feedback process would make it stronger for me, from a teacher perspective.

Researcher: To understand their needs maybe?

Rosy: Yes, exactly to find out what they want because you say focus on tense, articles, and relative clauses; I found sometimes they were speaking very well, using relatively complex language that I did not see that they were necessarily speaking badly sometimes. Yeah so having to try to come up with things on the spot to reformulate, to make it more complex, sometimes felt a bit artificial. It's like I am doing it unnecessarily but if that's their focus, if that's what they want to improve the most then I think I could have done more looking back to it.

Researcher: Were there instances where you felt that the students were trying to avoid using these forms?

Rosy: Sometimes they did not use them and I thought it still sounded fine because everything they were saying was appropriate and if they had used a lot of RC in some cases, I think it would have maybe sounded a bit forced? I did notice that they did use them sometimes and so I did not feel there was much need to reformulate when they've already demonstrated it. So I did not have a feeling that they were particularly avoiding those forms, but some of the most complex tenses were a bit confusing and occasionally like with uncountable/countable nouns.

Researcher: How did you find interacting with your partners?

I am not really put off by the fact that I was talking to complete strangers. I enjoy meeting new people and as a teacher I am used to constantly you know being with a new group of people. It is a bit unusual in the tasks that talk about things like personal values and family and friends.

Those were quite personal things and I am not very familiar with people from Syria so I did not know how conservative they would be about revealing you know details of their live or their feelings and I did not know how that would be for them in terms of the task. but I felt I am fine with it. I felt they both have really a good sense of humour and were quite open but it's true that there were some cultural differences that came out in the task when we had shared our opinions, maybe we did not always agree or did not always have the same kind of approach, but I thought there was no discomfort on my part. It seems very natural and they were very friendly people.

Researcher: Please tell me, how did you find the idea of beginning with text chat?

Rosy: I think text chat gave us a little time to get to know each other before we started having conversations where we have to be more careful about turn taking and the tasks became a little more personal as well when we started speaking I think. So yeah it gave us a little time to develop a bit of a rapport, so I get to know their language, their errors in a more objective way as a scene on the screen instead of hearing their voice but I don't know how it felt from their perspective maybe easier because they could see my feedback and then keep track of it more easily for themselves but I am not sure.

Researcher: I noticed that most of your feedback was explicit, could you please tell what made you choose this approach to address the learners' errors?

Rosy: Yeah, in terms of reformulating, sometimes I was just near and about part of their answer and try to add more complex phrases or some new vocabulary or suggest expressions that I thought would be appropriate for the scenario. I think that I did not want to talk down to them because they are adults, but I thought it's useful to highlight or raise their awareness of certain errors or certain points, but I wanted to get them speaking and the flow of their ideas going, but then I tried to be pretty quick about highlighting errors so that it was still fresh in their mind and they could immediately draw a connection. But I think sometimes I was trying to

reformulate with RCs and sort of hoping that they might repeat something similar but it did not always happen. I felt the feedback on new vocabulary was slightly better. I noticed some of them, there was a bit of feedback like they would hear me say something or suggest something and then a little while later I hear them say it correctly themselves. I like to be sensitive to the learners in terms of I want them to feel comfortable and that they have a good amount of time to express themselves. I don't think it's helpful to constantly interrupt but I think if you wait too long then I think it's harder for the learners to pick up on things and really make connections so I do like to be explicit but I also think it's good to get them speaking and I think the more they could speak, the more they produce, the better for their fluency and then I could get more examples of their language So I like to do a bit of both. I think there is a role for both and I think I still sometimes struggle with finding the perfect balance for each learner.

Researcher: OK

Rosy: During the text thing I felt it was a lot easier to manage the feedback. I did not feel it's bad interrupting them in their speech because they were typing anyway. So I felt I was giving more explicit feedback maybe when it was written, whereas with the spoken I think I was holding back a little bit more to not keep cutting them off to the explicit. So I think I was trying to be more implicit like with my responses, trying to give them new language and trying to reformulate slightly using more complex ways to say something. So in the spoken I was a little bit more aware of not, trying to not to be rude or not cut them off in the middle of saying something. It was nice to have the consistency; we were used to each other sort of style and voice maybe a little bit. I think maybe partly because I did not feel I knew them that well; I was trying to still get them to speak and feel at ease so I think there was more apprehension about interrupting them.

Researcher: OK. Please tell me how did you find the tasks?

Rosy: There was an interesting mixture, pretty diverse in terms of what we had to do. I think overall the tasks were pretty clear. I also think that with picture-based tasks, they were a little bit limited in terms of the kind of language that we could produce but obviously I see the logic in starting with simpler tasks and making them more complex as you go. The latter tasks were more challenging in terms of the language, but you know when you are working with adults I think it's ok to go into more interesting topics because then people would focus more on their thoughts and their response rather than trying to overly monitor their language, so maybe they are being more natural.

Researcher: How well did you feel you worked with your partners?

Rosy: It's hard to judge because I did not hear from the students directly but I don't know I think most of the time we were able to accomplish the goal of the task. There were however, one or two times where I felt that the students did not have as much to say about the topic where they were getting a little impatient, so I was feeling maybe I could have done more to help mitigate any sort of frustration or try to offer more help or more language but I felt overall we were able to accomplish the task and produce a decent amount of language.

Researcher: How effective do you think was the feedback that you gave to the learners?

Rosy: I thought maybe my feedback in the earlier tasks was maybe better because I could provide more feedback in the text, I felt. Not sure how the students felt. And then in the latter tasks maybe not as effective I think, particularly in relative clauses. It was really hard to get them to produce the forms and I could have maybe been more forceful about getting them to say it back to me or getting them to type it out. I felt in terms of offering new vocabulary and more complex expressions, they were repeating some or producing some, but I think in terms of producing RCs I think I could have done better.

Researcher: Did you notice any change in learners' performance throughout the sessions?

Rosy: To be honest, I don't have a strong impression there was any massive difference but I think with doing similar tasks over and over. It was easier to see some improvement in the text-based chat but again sort of each time the task was different so the type of language was a little bit different and with the voice chat perhaps with time I know also they would get a little bit more comfortable, but I can't recall specific examples of ways that I saw them significantly improving. I think the sessions were relatively short and spread apart so it's kind of hard to track. I wasn't keeping a very good track of the number of errors and number of times they were corrected, but I think from my memory, I think it was a little bit more evident improvement from the begging; from the first task to the last task.

Researcher: Is there anything else you would like to comment on?

Rosy: Nope, I think that was pretty it

Researcher: OK! Thank you very much for taking part.

Rosy: you are very welcome!

<div align="center">***</div>

Appendix G

Codes' Lists

a. Codes used for the analysis of SLs' planning notes

Language Aspects	Definition	Example
Grammar	Instances where the learners attended to the grammatical aspects of their production (e.g., sentence structure, tense, articles, prepositions).	e.g., The main problem that Linda has, is to find the time between her responsibilities. ***The learner then edited this sentence to be as follows:*** The main problem is that Lind has to find time for her different responsibilities.
Meaning	Instances where the learners wrote ideas related to the content of the task	e.g., On the shelves, sound system
Vocabulary	Instances where the learners listed single words; to refer to particular items they saw in a picture, for example.	e.g., bed Chairs Candles Pillow
Spelling	Instances where the learners kept editing a word till they figured out the right spelling	e.g., carbage (**edited to** garbage
pronunciation	Instances where the learners used an online dictionary to check how a particular word is pronounced	None

b. <u>Codes used for the analysis of SLs' L2 production</u>

Coding category		Description	Example
Complexity	Syntactic Complexity	Number of clauses per AS-units (Yuan & Ellis, 2003). Length of AS-units (i.e., Number of words per AS-unit) (Kawauchi, 2005).	Anas: Presumably the concert would be an instant satisfaction but a pain on the long run (1 AS-unit, two clauses, 15 words
	Complex grammatical structures	Use of relative clauses and other complex structures per task (Kawauchi, 2005)	Majd: the second patient <u>who</u> is an accomplished violinist and a 12 yo child should receive the heart because she's literally a child and she still hasn't lived that much
	Syntactic Variety	Total number of different grammatical verb forms used in the task in terms of tense, voice, and modality (Yuan & Ellis, 2003)	Majd: Well, I **think** the teacher **should receive** it because first he **has** two children that he **has** to take care of and **he is basically teaching** generations so he'**s** a treasure for sure In this example, the learner produced different grammatical verb form: Tense= simple present/ present progressive And 1 instance of using Modality "should receive"
	Lexical variety/richness Lexical accuracy/appropriacy	The total number of different words occurring in a text or utterance was divided by the total number of words (Yuan & Ellis, 2003). Learners might use a variety of words during a task; however, the words might not be appropriate for the task context (in terms of meaning).	**e.g.,** In the bottom right corner, there's a **garbage can,** which is full. Instead of saying: there's a **trash can/bin** which is full

Accuracy	Grammatical Accuracy	Number of error-free clauses per total number of clauses per task Number of errors as compared to the total number of words produced per task	e.g., If this **happen** to me, I will like prove myself. In this example, we have two clauses; only the second one is error-free whereas the first one has a subject-verb agreement problem.
Fluency	mean length of unfilled/silent pauses	total silence greater than 0.25	
	mean length of filled pauses	learners' using (erm, err)	
	Speech rate	mean number of meaningful syllables divided by speaking time, excluding, repetitions, false-starts.	
	Repair measures	mean number of false-starts, reformulations per minute	

c. Coding Categories of the type of feedback provided by NSs/STs

Category	Level	Examples
Language focus	Lexical Grammatical	SL: it just has a curved and straight lines on it NS: we would say 'curvy lines' or wavy lines' SL: and there's a chair in front on the computer desk NS: you mean, 'in front of'?
Type of feedback	Explicit Or Implicit (Recast) (Confirmation Check)	SL: Should I start telling you what I'm seeing in the picture? NS: yes please. Also, you would say 'what I see in the picture'. Present simple instead of present continuous. SL: on the left upper corner, NS: in the upper left corner? SL: in the down left corner, NS: you mean, 'in the bottom left corner'

d. Coding Categories for the Stimulated Recall Interviews.

Category	Definition	Example
1. Anxiety/stress	This happened quite often, and learners reported different reasons for feeling a bit anxious either before or during the online sessions	
a. Lack of practice/new experience		I was also anxious about the fact that it's been a while since I last write/speak in English, and the fact that the whole experience was new to me.
b. lack of confidence		As the sessions started, I was anxious about my ability to express my thoughts to my partner. I did not feel confident enough to write or speak in English at the beginning.
c. Chatting with a more proficient L2 speaker		I had to talk to a NS; someone who was totally a stranger for me, and I was very anxious about that.
d. Feedback delivered by their partners		She corrected me whenever I made an error or wrote an informal word. She was correcting my language all the time, so I felt anxious, stressed out, and tried to focus more on my language to avoid making more errors.

Rapport between the participants		My partner was very friendly and supportive. Like you know we understood each other very quickly, and we shared similar thoughts on most of the topics that we worked on. Even though we have not met before, I think we got on well together.
Silence/hesitation during the online interaction:	Instances where the learners were pausing or seemed a bit hesitant to write/speak	As I've been studying French for 3 years now, it's been a while since I last practiced speaking/writing in English. That's why I was confused about using (spelling/pronouncing) some words.
Language interference		I also faced some challenges when writing, influenced by my French studies, and the fact that I was also trying to learn Turkish (enrolled in a Turkish course). You could imagine how the three languages were mixed up in my mind as each language has different rules in terms of grammar, sentence structure, and so on. I should admit that it took me some time (the first couple of sessions) to switch back to the English mode. This might justify the delay in my responses and why I was hesitant sometimes when typing my messages; I felt like I needed more time to process the language and form an appropriate answer due to the interruption that I had while learning other languages.

Focusing on multiple aspects of their production		I am that kind of a person who prefer to deliver meaning in the most appropriate way in terms of grammar, structure, spelling, etc.
Thinking of the right vocab/word to use		I already had an answer in my mind but I was trying to look for the right expression/word to use in that context.
Mode of interaction	The lack of thinking time during the voice chat as opposed to text chat was making the learners pause more often.	I took some time to think of what to say and how to say it.
Focus during the sessions: Form		The fact that I was interacting with a NS made me more attentive to the language I produced; I was trying to make as few errors as possible.
Word choice		This also greatly affected my vocabulary choices during the sessions.
Form & meaning		I had to think of how to express my thoughts, put them in words, and produce grammatically correct sentences, so that my partner could easily understand what I was saying.
Advantages of text chat a. less threatening		I found text-based chatting very efficient and less threatening (compared to F2F or voice chat) when it comes to talking for the first time to a complete stranger using another language.

Drawbacks (overlap in turn-taking)	Happens when both interlocutors type at the same time, addressing two different points. This often cause interruption.	There was one minor problem though; during the text-based chat, there were times when I wrote a message to my partner but then noticed that she was asking me a question about something different. In this case most of the time, I felt that it's better to delete what I wrote and answer the question first to maintain the flow of the conversation.
Perceptions towards the planning time (Practical/helped lowering the stress level) helped learners focus on different aspects of production		I liked the idea of notetaking very much; it was a very practical and stress reliever strategy. You know like I already had something in mind to talk about, and consequently I managed to focus on other aspects of my language (spelling, grammar, etc) during the session.

Perceptions towards the online sessions (focused)		I found the sessions very focused as we had to spend a particular amount of time talking about one topic, trying to find solutions and thinking of all the pros and cons of these solutions before making a decision.
(a secured learning context)		I tired online platforms before and I found it very difficult to continue as not all the people I talked to were serious about learning another language. Whereas during this experience, I felt safe because I knew that this is a real project which is approved by a well-known university.
(unforgettable)		the whole online experience was unforgettable.

Perceptions towards the tasks' type (sequenced)		What I liked the most about the tasks was the fact that they were sequenced based on their complexity (simple-complex). This was great as I believe that it would be very distressing to start working on complex tasks and discussing real problems with someone you've never met before. I think I needed some time to get ready to open up and share my thoughts regarding particular topics with my partner.
(diverse)		At the same time, I liked the variety of the task's types we worked on; I was excited to know what kind of tasks we'll be working on each week. I assume it would have been very tedious to work on the same type of tasks during all the sessions.
Perceptions of the tasks' content (resemble real-life situations)		All the topics were taken from everyday situations and so you get a great opportunity to improve your speaking/writing skills as well as expand your knowledge and vocabulary items regarding these topics.
Attitudes towards the feedback delivered by their partners. (noticed)		I admired the way she corrected my language; she was very considerate, corrected my in an indirect way so that I don't feel intimidated or something.

Appendix H:
Pilotees' background information

A. **SLs pilotees' background information**

Syrian learners (SLs)	Age	Years of studying English as a foreign language	Level of study
Kareem	25	17 years	Graduate
Asma	24	18 years	Undergraduate
Doaa	23	17 years	Undergraduate
Obada	27	18 years	Graduate
Eiad	28	19 years	Graduate
Jana	21	15 years	Undergraduate

B. **NSs/STs pilotees' background information**

NSs/STs	L1 Status	Age	Years of teaching English
Alma	ST	31	3 years
Tamara	ST	30	2 years
Sophia	NS	29	4 years