Ghadah Albarqi*

Investigating EFL oral production in a technology mediated TBLT context

https://doi.org/10.1515/jccall-2023-0030
Received December 12, 2023; accepted May 21, 2024; published online July 15, 2024

Abstract: In the rapidly expanding domain of online learning, investigating the effectiveness of technology-mediated task-based language teaching (TBLT) in improving English as a Foreign Language (EFL) oral production becomes paramount. Therefore, the present study was conducted over two months to examine how EFL learners' oral production, measured by complexity, accuracy, and fluency (CAF), changed throughout technology-mediated TBLT instructions. This study also examined how the relationships among CAF aspects evolved. A cohort of 38 EFL learners underwent a series of oral tasks at the beginning and end of the study period, and repeated measures and correlation analyses were employed to analyse the data. The findings showed significant improvement in EFL learners' speech rate and accuracy of their utterances after two months. Moreover, the data indicated that fluency features were intricately connected with other CAF dimensions, whereas accuracy, lexical diversity, and syntactic complexity exhibited independent development. The identification of long-term relationships among certain CAF aspects provided valuable insights into the underlying mechanisms of EFL speech production. The results of this study have significant implications for both research and practice within the domain of technology-mediated TBLT.

Keywords: technology-mediated TBLT; online instruction; complexity; accuracy; fluency

1 Introduction

Task-based language teaching (TBLT) has emerged as a prominent pedagogical approach in the domain of communicative language teaching. Its primary objective is to create an authentic learning environment that facilitates real-time communication and the development of learners' communicative skills (Faez and

***Corresponding author: Ghadah Albarqi**, Foreign Language Department, Taif University, P.O. Box: 11099, Taif, Saudi Arabia, E-mail: gadah.g@tu.edu.sa. https://orcid.org/0000-0001-9043-5881

Open Access. © 2024 the author(s), published by De Gruyter and FLTRP on behalf of BFSU. This work is licensed under the Creative Commons Attribution 4.0 International License.

Tavakoli 2019; Robinson 2011; Skehan 1998, 2014). At the core of TBLT is the notion of a task, representing a language activity that focuses on meaning, involving a specific gap or problem. Learners are required to use appropriate linguistic resources to effectively communicate and achieve a desired outcome (Bygate 2016; Ellis 2003, 2009, 2018). The theoretical foundations of TBLT draw on the interaction hypothesis (Long 1996), emphasizing the importance of negotiating meaning in improving L2 speaking. Additionally, the output hypothesis (Swain 2005) plays a role in TBLT by encouraging language output opportunities for learners to identify their gaps in language learning and reflect on their learning progress.

Similar to the concept of TBLT, computer-assisted language learning (CALL) has experienced significant growth. Despite the widespread integration of technology in language classrooms and its recognized developmental benefits (Baralt and Gómez 2017; Chong and Reinders 2020), researchers emphasize the necessity for a more structured and theoretically grounded approach to CALL (Chong and Reinders 2020; González-Lloret and Ortega 2014; Ziegler 2016). This is crucial for designing computer-based activities that are pedagogically effective. Responding to this need, researchers have begun exploring the possibility of merging these TBLT and CALL (González-Lloret and Ortega 2014; Ziegler 2016) to advance the development of both fields and leverage their synergies.

The convergence of TBLT and technology has given rise to a promising approach known as technology-mediated TBLT (see González-Lloret and Ortega 2014; Ziegler 2016). Recent years have witnessed significant developments in technology-mediated TBLT, exploring the potential of technology in addressing perceived challenges associated with TBLT, such as limited opportunities for authentic communication (Carless 2012; Smith and González-Lloret 2021). The issue of limited authentic communication is especially prominent in EFL contexts, where students have limited opportunities to practice the target language beyond the classroom (Albarqi 2020; Chen 2021). This scarcity often leads EFL students to hesitate to speak during L2 classes (Albarqi 2020), necessitating the design of an L2 learning setting that mirrors real-life contexts. In this study, technology was aligned with TBLT principles to optimize the teaching and learning experience. The present study investigated the development of L2 oral language production, defined in terms of Complexity, Accuracy, and Fluency (CAF) triad. Throughout the study, the terms “oral production”, “oral performance” and “speech production” are used interchangeably. This study aims to gain insights into the development of EFL oral performance and the changes in the relationships among CAF aspects within this technology-mediated TBLT setting. In this context, the integration of TBLT into online instruction via a Learning Management System (LMS) is defined as technology-mediated TBLT instruction.

2 Literature review

2.1 Technology-mediated TBLT and oral production

With the rapid expansion of online learning and teaching, a growing body of research is aimed at exploring the efficacy of technology-mediated TBLT on L2 proficiency and production. Most extant research on the impact of technology-mediated TBLT on the development of CAF aspects has predominantly focused on written contexts (e.g., Hokamura 2018; Kusyk 2017), with a scant investigation into the longitudinal development of CAF in oral modality. The majority of existing studies have employed a cross-sectional approach, investigating the effectiveness of aligning various task features, such as task complexity and sequencing (e.g., Baralt 2014), or planning (Hsu 2015; Park 2010) on L2 oral production.

An early study on the longitudinal development of oral production by Abrams (2003) spanned a semester comparing synchronous and asynchronous computer-mediated communication in terms of promoting oral skills, specifically focusing on the number of words, lexical richness and diversity, and syntactic complexity. While the synchronous group indicated a noteworthy increase in the quantity of language (number of words) produced, no significant differences in lexical or syntactical complexity emerged. However, this study did not investigate other dimensions of CAF, such as fluency and accuracy. Another study conducted by Volle (2005) examined the effects of online distance instruction on L2 learners' oral production in terms of articulation, accuracy, and proficiency scores over a semester. The study involved four activities in which participants engaged in voice conferencing and indicated a substantial improvement in the learners' oral proficiency scores.

Satar and Özdener (2008) examined the effects of online distance instruction on L2 learners' oral production during four weeks of computer-mediated communication and found that text chat and voice chat groups improved students' speaking proficiency, with text chat proving as effective as voice chat for beginner-level students. In another study, Van der Zwaard and Bannink (2016) investigated the occurrence of negotiation of meaning in a computer-mediated TBLT environment. The participants included non-native speakers and native speakers of English who were assigned two different tasks including a culturally specific activity and a collaborative decision-making exercise designed to stimulate negotiation of meaning. The results indicated that non-native speakers frequently refrain from engaging in the negotiation of meaning despite encountering comprehension difficulties. The study concluded that neglecting the negotiation of meaning in digital task-based language teaching can have detrimental effects

on task performance, outcomes, and evaluation. Li and Zhang (2023) investigated the accuracy and fluency in speech of 45 Chinese undergraduates over a semester of online teaching, showing a noteworthy improvement in accuracy but no significant enhancement in fluency. Notably, this study did not align their instructions with the TBLT approach.

In summary, exploring the existing literature in this field reveals certain limitations in specific research areas. In particular, some studies, such as those by Abrams (2003), Li and Zhang (2023), and Volle (2005) did not explicitly incorporate TBLT principles in their instructional design or did not report their use of TBLT in their studies. In addition, some of the previous studies, including those by Abrams (2003), focused on language learning within blended classes rather than a purely online environment introducing additional factors and variables that may impact language learning outcomes. The integration of online and face-to-face components may create a different learning environment compared to an exclusive online setting. In addition, within technology-mediated TBLT studies, attention has primarily been on individual aspects of the CAF triad (e.g., Abrams 2003; Van der Zwaard and Bannink 2016; Volle 2005). Although these focused investigations offer valuable insights, a comprehensive understanding of changes in L2 processing and production necessitates an examination of various CAF dimensions (Robinson 2007; Skehan 1998; Van Moere 2012).

2.2 Development and relationship among CAF aspects

Within TBLT, scholars have probed into the changes in L2 processing and production by examining the CAF triad (see Robinson 2007; Skehan 1998). Notably, the CAF aspects mark the fundamental stages of development in the L2 linguistic system (Housen et al. 2012), serving as sensitive indicators of L2 processing and production (Pallotti 2009; Robinson 2007; Skehan 1998; Van Moere 2012; Vercellotti 2017) as well as predictors of L2 oral proficiency (Tavakoli and Skehan 2005). The rationale for employing CAF in L2 research is "to account for how and why language competencies develop for specific learners and target languages, in response to particular tasks, teaching, and other stimuli, and mapped against the details of developmental rate, route, and ultimate outcomes" (Norris and Ortega 2009, p. 557).

L2 researchers often aligned the CAF triad with the major components of Levelt's (1989) model of L1 speech production (Robinson 1997, 2001; Skehan 1998; Skehan et al. 2016). Levelt's (1989) model commences with the Conceptualizer, which is responsible for generating preverbal messages. This is followed by the Formulator stage, where grammatical encoding takes place, including aspects like lexical access as well as syntactic and phonological encoding. The process culminates with the

Articulator producing the speech while the Monitor inspects utterances before and after articulation (Levelt 1989). Associating the CAF dimensions with Levelt's model is significant due to its reputation as a robust and validated framework for understanding speech production (Skehan 2009). Despite ongoing debates about the definition and operationalization of certain aspects, particularly fluency, this study incorporated the commonly employed definition in L2 research.

Firstly, complexity refers to the linguistic properties of L2 utterances, encompassing a diverse vocabulary, varied sentence structures, and the ability to convey complex ideas effectively (see Housen et al. 2012). According to Housen et al. (2012), complexity development entails the assimilation of new L2 elements. Accuracy, on the other hand, refers to the correctness with which a speaker uses language, encompassing grammatical correctness, appropriate word choice, and adherence to the rules and conventions of the target language (Housen et al. 2012). Improvement in accuracy is considered an indicator of efficient processing at the Formulator and Monitor stage (Albarqi and Tavakoli 2023; Kormos 2000, 2006; Wang 2014).

Fluency pertains to the smoothness and the natural flow of speech (Segalowitz 2010), involving the ability to speak effortlessly with control over various aspects of speech production such as speed, hesitation, and pauses (Lambert and Kormos 2014; Segalowitz 2010; Tavakoli 2019). Furthermore, an essential component of fluency is the speakers' ability to manage communication breakdowns and repairs (Segalowitz 2010; Tavakoli 2011). Although fluency received less agreement in L2 literature, the framework proposed by Tavakoli and Skehan (2005) seems plausible and has been validated in several TBLT studies (e.g., Garcia-Ponce and Tavakoli 2022; Kuiken and Vedder 2008; Lambert and Kormos 2014; Suzuki and Kormos 2023). This framework categorizes fluency into three aspects: repair, including repetitions, replacements, reformulations, and false starts; breakdown, comprising silent and filled pauses; and speed. Research has indicated that the efficient production of fluent speech is likely tied to proficient processing at both the Formulator and Articulator stages (Segalowitz 2010; Tavakoli 2019). Importantly, the three dimensions of CAF are considered independent constructs (Leonard and Shea 2017; Norris and Ortega 2009) and might not develop linearly or at the same rate (Spoelman and Verspoor 2010).

Research on CAF has generally employed cross-sectional designs (e.g., Foster and Tavakoli 2009; Garcia-Ponce and Tavakoli 2022; Kim and Tracy-Ventura 2013; Lambert and Kormos 2014; Tavakoli and Foster 2011). Exploring the longitudinal development of CAF in speaking, research generally indicates improvement in CAF features in instructed settings over time (e.g., Ferrari 2012; Tonkyn 2012; Vercellotti 2017; Yu and Lowie 2020). However, the majority of these studies have not explored the long-term relationships among CAF constructions over an extended period. The existing investigations of the CAF measures' interrelationships have shown

mixed results (e.g., Leonard and Shea 2017; McManus et al. 2021; Vercellotti 2017; Yu and Lowie 2020). Leonard and Shea (2017), for instance, found CAF interrelationships in the pre-test phase, but not in the post-test, while McManus et al. (2021) identified persistent ties between fluency and lexical elements. Vercellotti (2017) investigated the relationship among these CAF components through correlation analysis showing that there were linear changes in CAF aspects. Notably, the results indicated that all CAF constructs exhibited positive correlations and concurrent growth with no trade-off effects (i.e., focusing on one CAF aspect has a detrimental effect on another) in CAF. This revealed that students did not exclusively concentrate on enhancing one CAF construct at the cost of another. Yu and Lowie (2020) investigated the accuracy and complexity of the speech of 10 EFL Chinese undergraduates over four months, suggesting a noteworthy shift in the relationship between complexity and accuracy, transitioning from a negative to a positive correlation. This transformation suggested a shift from a competitive to a supportive relationship among these aspects during the later stages of development (Yu and Lowie 2020).

In short, the research on CAF has primarily employed cross-sectional designs, which do not provide insights into the evolution of CAF. Furthermore, inconsistent results arise when exploring the long-term relationship among CAF features. Besides, this area of investigation has rarely been examined in the technology-mediated TBLT context. To address these gaps, this study aims to investigate the development of EFL oral production, as represented by CAF, within a technology-mediated TBLT framework. It further seeks to shed light on how the CAF aspects interplay over time.

3 Methodology

3.1 Research questions

The present study was particularly motivated by a desire to document how EFL learners' oral production (as measured through CAF) developed throughout a semester of exclusive online instruction aligned with the TBLT approach amidst the COVID-19 era. Additionally, this research aimed to study how the relationships within the CAF triad evolved over this period. The research questions (RQs) addressed in this study are as follows:

RQ1: Does the oral performance of EFL learners, measured by CAF, change over a semester of technology-mediated TBLT instruction?

RQ2: How do the relationships among CAF dimensions change over a semester of technology-mediated TBLT instruction?

3.2 Methods

The study employs a pre-test and post-test design to investigate how EFL speech production and the relationship among CAF change for two months of online instruction aligned with TBLT. The analysis focuses on measures of CAF derived from the speech recordings, as elaborated in Section 8.

3.2.1 Participants

A cohort of 46 undergraduate students from the Information Technology Department of a state university in Saudi Arabia participated in this study. All participants completed at least 7 years of English language education at primary, intermediate, and secondary schools. Eight students were excluded because they could not complete all the requirements of the study. Informed consents were obtained from all participants indicating their understanding of the research's ethical considerations and their willingness to participate in the study. Participants' proficiency levels were assessed using the Oxford Placement Test (OPT); 21 students demonstrated at the A1 level, 14 at the A2 level, and 3 at the B1 level, according to the Common European Framework of Reference for Languages (CEFR) framework.

3.2.2 Tasks

The pre-test and post-test procedures in this study involved the completion of three tasks by the participants at the beginning and end of the semester. These tasks were adapted from previous literature (Albarqi and Tavakoli 2023; Heaton 1966) and included a personal information task, a football narrative prompt, and a museum narrative prompt (refer to Appendix B for details). The personal information entailed speaking about the weekend routine. The football task required describing a story about a group of boys playing football, wherein the ball fell into a hole, and one of the boys devised a plan to get it back. The museum narrative task recounted a story of a class going on a trip to a museum, wherein a fire broke out and students were eventually rescued. Tasks that involve picture-based narrative and personal information are commonly employed in L2 classes and are viewed as ecologically valid in L2 research (Garcia-Ponce and Tavakoli 2022; Préfontaine and Kormos 2016; Tavakoli and Foster 2011). Likewise, monologic tasks have been frequently employed in longitudinal studies (e.g., Li and Zhang 2023; Vercellotti 2017; Yu and Lowie 2020). All instructions were given in Arabic, participants' L1, and were counterbalanced and presented to participants through the LMS.

3.3 Implementation of technology-mediated TBLT instruction

The study was conducted online during the Covid-19 pandemic as part of a larger project. The participants were enrolled in an Intensive English for Academic Purposes (IEAP1) program, which entailed 12 h of weekly exclusive online instruction aimed at enhancing language skills in speaking, writing, reading, and listening. Each language skill received 3 h of instruction within the program. The course utilized the textbook *Unlock Level 2* (Dimond-Bayir et al. 2019; Westbrook et al. 2019). The online course was implemented and delivered by the researcher. The task phases, implemented through the LMS, were designed following the TBLT principles (see Ellis 2003, 2008, 2009; Ellis et al. 2019; Faez and Tavakoli 2019; Long 2015; Willis 1996; Willis and Willis 2007) (see Table 1 and Appendix B). Previous L2 research indicates that these phases can influence CAF production in several ways. For instance, the introduction of pre-task planning has been found to promote fluency and complexity (Ellis 2009). Pre-task activities were designed to familiarize learners with the topic by providing background information and necessary vocabulary (see Ellis 2003). Research indicates that tasks involving familiar content have the potential to enhance both fluency and accuracy (Skehan 2009). However, there is ongoing debate regarding the optimal duration for pre-task planning. While 10 min is commonly suggested as the default time, some argue that 5 min is the ideal duration (Ellis et al. 2019; Faez and Tavakoli 2019). In the current study, the pre-task phase was conducted using diverse technological tools, as outlined in Table 1 and Appendix B. This study was implemented via Blackboard Collaborate Ultra (version 3800.0.6.rel.12 + 96497c7) which provides both audio and video conferencing tools.

The on-task phase entails working in small groups within breakout groups on the Blackboard (audio conferencing tool is used in breakout groups for practical reasons). During this phase, students are encouraged to actively participate in problem-solving, reasoning, and expressing opinions, agreements and disagreements. This participation fosters the production of complex structures such as subordinate clauses, ultimately contributing to a higher rate of syntactic complexity in L2 speech (see Robinson 2007).

Providing ample planning time during task performance can encourage learners to attend to both form and meaning, thereby improving both fluency (Faez and Tavakoli 2019) and accuracy (Wang 2014).

During the post-task phase, also known as the language focus phase, according to Willis (1996), the instructor plays a role in pinpointing language errors, providing feedback and clarifications, and guiding students in the application of appropriate grammar and vocabulary. Some researchers have noted that incorporating task

Table 1: Online instruction aligned with TBLT principles.

Phases	Activities	Aim	Technological tools
Pre-task phase	– The instructor introduces necessary vocabulary and phrases to prepare students for task performance (but they are not required to use the exact words) – Pre-task planning (5–10 min) can be used to create an outline of the task (e.g., note-taking, diagram, chart, table, idea map)	– Making the tasks more engaging and relevant – Motivating learners – Preparing them for the task – Familiarizing students with the topic and content (Ellis et al. 2019) – Easing pressure on the cognitive resources (see Ellis et al. 2019; Faez and Tavakoli 2019)	eBook (presentation plus tool) LMS class collaborate ultra Videos Slides Breakout groups Interactive whiteboard
On-task phase (three stages)	**1-Task:** the students perform the task within breakout groups. The instructor visits these groups to check their work and provide support and clarification **2-Planning:** the students prepare to present their work **3-Report:** the instructor ends the breakout groups, and students return to the main room. They share their screen and present their work to the whole class. Supporting materials during the task performance is desirable (e.g., video, image, planned notes, diagram, idea map)	– Encouraging collaborative work – Utilizing available resources to accomplish a desired outcome – The focus on content and task completion – Students need to employ problem-solving skills to complete the task – Within-task planning can be provided to encourage learners to pay attention to form and meaning (Faez and Tavakoli 2019) – Supporting materials during the report stage reduce pressure on the memory (Faez and Tavakoli 2019)	Class collaborate ultra Breakout groups Padlet Slides Microphone Sharing screen
Post-task phase	– Students can compare their work to that of other groups – Students evaluate their peers' performance – Students reflect on their performance – The instructor provides feedback in oral and written modes – The instructor presents a grammar explanation – The instructor introduces a survey or quiz to highlight certain form – Students repeat the steps with a similar task	– Evaluating task performance – Providing feedback – Highlighting common errors – Presenting grammar explanation – Promoting fluency, lexical diversity, and syntactic complexity through task repetition – Promoting accuracy by highlighting errors	Class collaborate ultra Poll tool, mentimeter Breakout groups Chatbox Interactive whiteboard

repetition in the post-task phase can augment fluency, lexical diversity and syntactic complexity (De Jong and Perfetti 2011). Additionally, procedural repetition, where students apply similar steps with different tasks, has been found to enhance accuracy (Kim and Tracy-Ventura 2013; Patanasorn 2010) and lexical complexity (Kim and Tracy-Ventura 2013). TBLT instruction diverges from traditional classroom approaches in its emphasis on meaning. In TBLT, learners primarily concentrate on the content, encompassing semantic and pragmatic meaning, rather than being focused on linguistic form. The linguistic form is introduced in the final stage, known as the post-task phase (see Ellis et al. 2019; Faez and Tavakoli 2019). This approach reflects a shift in language pedagogy towards a more communicative and meaning-oriented method.

3.4 Data collection

Data collection was conducted in two phases, at the beginning and end of the semester. The participants were required to complete three oral tasks at the beginning and end of the semester (see Appendix A), with a time interval of 8 weeks (around 2 months) between the two phases. Several methodological issues arose during data collection such as technical problems, network coverage, and audio quality. These issues are typical during online data collection (see Hampel and Hauck 2004; Wang 2004).

3.5 Coding CAF measures

The participants were instructed to speak for 1 min in each task, totalling approximately 228 min (equivalent to nearly 4 h) across the two phases. After transcribing the data, it was segmented into AS-units, following guidelines introduced by Foster et al. (2000). Subsequently, the data underwent coding for various CAF measures, using procedures outlined in previous L2 studies (such as Ellis and Barkhuizen 2005; Foster and Tavakoli 2009; Garcia-Ponce and Tavakoli 2022; Tavakoli and Skehan 2005) (see Table 2). Fluency includes three dimensions: speed, repair, and breakdown (silent/filled pauses) as described by Foster and Tavakoli (2009). PRAAT was used to calculate silent pauses (Boersma and Weenink 2008). To ensure coding reliability, a second rater coded 10 % of the data for interrater reliability. The results showed a high level of agreement, with 98 % agreement for the type-token ratio (lexical diversity), 97 % for syllables per minute (speed fluency), 95 % for syntactic complexity (mean length of utterance), 87 % for error-free clauses (accuracy), 89 % for repair, and 92 % for pauses (silent and filled pauses). Disagreements were resolved through discussion between the raters.

Table 2: CAF measures.

CAF	Measures	Definition
Syntactic complexity	Mean length of utterance (MLU)	The text was pruned to remove all fillers. Then, the number of words was counted and divided by the number of AS units (see Foster and Tavakoli 2009).
Lexical complexity	Type token ratio (TTR)	The type-token ratio was calculated using Coh-Metrix. A higher TTR suggests a wider variety of words was used (see Foster and Tavakoli 2009).
Accuracy	Percentage of error-free clauses (PErFrC)	This involved manually counting the error-free clauses and dividing them by the total number of clauses, then multiplying the result by 100 (Ellis and Barkhuizen 2005; Tavakoli and Skehan 2005).
Fluency (breakdown)	Silent pauses	The number of silent pauses over 0.25 s in a speech sample was divided by the total time of the speech in seconds and multiplied by 60 (see Kormos and Dénes 2004).
Fluency (breakdown)	Filled pauses	The frequency of filled pauses (uh, umm, err) in the speech sample was divided by the time of the speech in seconds and then multiplied by 60 (see Kormos and Dénes 2004).
Fluency (repair)	Repair	The total number of repairs (reformulation, repetition, false starts) in the speech sample.
Fluency (speed)	Syllables per minute (SpM)	Number of syllables per minute was calculated in an unpruned text, using a website: https://syllablecounter.net/count

4 Results

Descriptive statistics of CAF measures in the two phases are presented in Table 3. Most CAF measures indicate an improvement from phase 1 to phase 2 as illustrated in Figure 1.

To address the first research question, a repeated measures analysis of variance was employed. Significant results were interpreted based on Cohen's (1988) guidelines for the partial eta squared values: 0.01 indicated a small effect size, 0.06 indicated a moderate effect size, and 0.14 is a large effect size.

The first research question aimed to determine whether there were differences in the CAF-based oral performance in the two phases. Table 4 demonstrated a significant change (Wilks' Lambda = 0.624; $F = 2.67$, $p = 0.028$; $\eta^2 = 0.376$) indicating changes in CAF features after the period of online instructions aligned with TBLT.

Within-subjects analysis of variance indicated significant differences in two of the measures analysed, as illustrated in Table 5. First, there was a statistically

Table 3: Descriptive statistics of CAF measures in the two phases.

Measures	Phase 1 *M* (SD)	Phase 2 *M* (SD)
MLU	33.97 (6.53)	35.78 (6.66)
TTR	1.72 (0.21)	1.71 (0.20)
PErFrC	110 (44.87)	126 (49.46)
Silent pauses	66.88 (11.46)	67.61 (13.15)
Filled pauses	52.81 (20.83)	50.39 (22.82)
Repair	11.26 (6.48)	12.1 (7.41)
SpM	368 (97)	405 (96)

Note: MLU = mean length of utterance, TTR = type-token ratio, PErFrC = percentage of error-free clauses, SpM = syllable per minute.

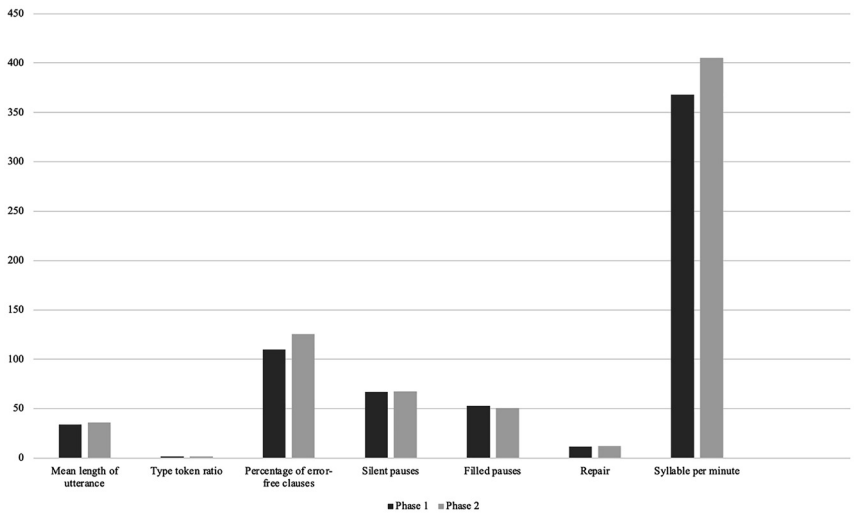


Figure 1: Means of CAF measures in the two phases.

Table 4: Results of repeated measures analysis of variance.

Effect	Wilks' lambda value	*F*	Sig.	Partial eta squared
CAF	0.624	2.67	0.028*	0.376

*$p < 0.05$.

significant increase in accuracy, measured as the percentage of error-free clauses, from phase 1 (M = 110, SD = 44.8) to phase 2 (M = 126, SD = 49.46), $p < 0.015$, $\eta^2 = 0.150$. This means that after two months, students' oral performance became significantly more accurate as measured by the percentage of error-free clauses.

Table 5: Within-subject variances of CAF measures in the two phases.

Measures	Mean difference (phase 1–phase 2)	*F*	*P*	Partial eta squared
MLU	1.81	2.505	0.122	–
TTR	0.014	0.253	0.618	–
PErFrC	15.922	6.55	0.015*	0.150
Silent pauses	2.148	0.166	0.350	–
Filled pauses	0.734	0.896	0.686	–
Repair	0.813	3.3	0.303	–
SpM	37.246	14.524	0.001*	0.282

*The mean difference is significant at the 0.05 level. MLU = mean length of utterance, TTR = type-token ratio, PErFrC = percentage of error-free clauses, SpM = syllable per minute.

The partial eta squared statistics (0.150) indicated a large effect size. In addition, there was a significant increase in syllables per minute in phase 2 (M = 405, SD = 96) as compared to phase 1 (M = 368, SD = 97), with $p < 0.001$ and $\eta^2 = 0.282$. The partial eta squared associated with this value is large (0.282) according to Cohen's (1988) guidelines. This suggests that students' oral performance significantly improved in terms of speed, as measured by syllables per minute. Taken together, the results of the first research question indicate that notable variations were observed in students' oral performance as assessed by CAF. Specifically, their speech showed a significant improvement in both speed and accuracy by the end of the two-month online instruction period indicating substantial development in their ability to articulate more quickly and accurately in English.

The second research questions aimed to explore how the relationships among the CAF aspects changed across the two phases. To answer this, a Pearson product-moment correlation analysis was conducted. Following Cohen's (1988) guidelines: correlations of r = 0.10–0.29 are seen as small, r = 0.30–0.49 as medium, and r = 0.50–1.0 as large. Tables 6 and 7 reveal two emerging patterns of relationships: the relationships among CAF dimensions on the one hand, and the relationships among fluency features on the other. Table 6 shows no significant correlations between accuracy (PErFrC), lexical diversity (TTR), and syntactic complexity (MLU), suggesting these dimensions were independent in the initial data collection phase. In contrast, fluency features correlated with other CAF dimensions. Specifically, there was a strong significant positive correlation between syntactic complexity (MLU) and syllables per minute (r = 0.42, p = 0.010). Syntactic complexity also correlated with silent pauses (r = 0.38, p = 0.019). This suggests that a more syntactically complex utterance might be produced faster and include silent pauses (further discussion on this will be presented in the next section).

Furthermore, there were strong significant negative correlations between lexical diversity (measured in TTR) and silent pauses (r = –0.43, p = 0.008), repair (r = –0.56, p = 0.000), and syllables per minute (r = –0.75, p = 0.000). This implies that EFL learners who use broader lexical elements in their speech tend to speak more

Table 6: Pearson product-moment correlation: CAF measures in phase 1.

Measures		MLU	TTR	PErFrC	Filled pauses	Silent pauses	Repair	SpM
MLU	Pearson	1	−0.249	−0.011	−0.255	0.377*	0.167	0.415**
	correlation		0.132	0.946	0.123	0.019	0.316	0.010
TTR	Sig. (2-tailed)		1	0.191	−0.208	−0.426**	−0.557**	−0.747**
	N = 38			0.251	0.210	0.008	0.000	0.000
PErFrC	*r*			1	−0.345*	0.102	−0.293	0.090
	p				0.034	0.541	0.075	0.589
Filled pauses					1	0.070	0.289	0.078
						0.677	0.079	0.640
Silent pauses						1	0.370*	0.402*
							0.022	0.012
Repair							1	0.326*
								0.046
SpM								1

Note: *Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant a the 0.01 level (2-tailed). MLU = mean length of utterance, TTR = type-token ratio, PErFrC = percentage of error-free clauses, SpM = syllable per minute.

slowly and with fewer silent pauses and self-repairs. Accuracy negatively correlated with filled pauses (r = −0.35, p = 0.034), indicating that EFL students who speak more accurately tend to pause less frequently. Regarding the relationships among fluency features, repair correlated with silent pauses (r = 0.37, p = 0.022), and syllables per minute correlated with both silent pauses (r = 0.40, p = 0.012) and repair (r = 0.33, p = 0.046). This suggests that EFL students who produce more repairs also make more silent pauses. Additionally, speaking faster is associated with more frequent silent pauses and repairs. Based on Cohen's (1988) guidelines, the correlations found among CAF in the first phase range from medium to large.

Table 7: Pearson product-moment correlation: CAF measures in phase 2.

		MLU	TTR	PErFrC	Filled pauses	Silent pauses	Repair	SpM
MLU	Pearson	1	−0.021	−0.261	0.138	0.308	0.389*	0.195
	correlation		0.900	0.133	0.408	0.060	0.016	0.240
TTR	Sig. (2-tailed)		1	−0.009	−0.137	−0.259	−0.560**	−0.585
	N = 38			0.957	0.411	0.116	0.000	0.000
PErFrC	*r*			1	−0.349*	0.138	−0.366*	0.367*
	p				0.032	0.408	0.024	0.023
Filled					1	0.035	0.534**	0.041
pauses						0.833	0.001	0.809
Silent						1	0.313	0.325*
pauses							0.055	0.046
Repair							1	0.310
								0.058
SpM								1

Note: *Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed). MLU = mean length of utterance, TTR = type-token ratio, PErFrC = percentage of error-free clauses, SpM = syllable per minute.

Table 7 presents correlations among CAF in the second phase. As with Phase 1, syntactic complexity, lexical diversity, and accuracy showed no significant correlations, suggesting that these dimensions may develop independently. However, fluency features correlated with other dimensions. Specifically, there was a weak positive correlation between syntactic complexity (MLU) and repair ($r = -0.39$, $p = 0.016$), implying that utterances with greater syntactic complexity are associated with more frequent repairs. Additionally, there was a strong negative correlation between lexical diversity (TTR) and repair ($r = -0.56$, $p = 0.000$), indicating that a broader lexical range in EFL speech corresponds with less frequent repairs. This relationship between lexical diversity and repair is consistent across both phases, as evidenced by Tables 6 and 7. Accuracy (PErFrC) demonstrated a significant negative correlation with filled pauses ($r = -0.35$, $p = 0.032$), suggesting that EFL students producing more accurate speech have fewer filled pauses. This trend remains stable over time. Moreover, in the second phase, accuracy also showed a negative correlation with repair ($r = -0.37$, $p = 0.024$) and a positive correlation with syllables per minute ($r = 0.37$, $p = 0.023$). This indicates that more accurate speech in the second phase is delivered faster with fewer repairs and filled pauses.

Regarding the relationships among fluency features in the second phase, a strong positive correlation exists between repair and filled pauses. Interestingly, in the first phase, repair correlated with silent pauses, suggesting that repairs can be linked with either filled or silent pauses. Additionally, there is a consistent positive correlation over time between syllables per minute and silent pauses. In summary, the findings from Tables 6 and 7 highlight that accuracy, lexical diversity, and syntactic complexity may develop independently. Conversely, fluency features show an interrelationship with other CAF dimensions. Notably, certain relationships are consistent across both phases: accuracy is linked with fewer filled pauses; lexical diversity correlates with fewer repairs and repairs are associated with either filled or silent pauses. Furthermore, syllables per minute consistently correlate with silent pauses. These relationships will be explained in the existing literature in the next section.

5 Discussion

This study aimed to investigate how oral production, as measured by CAF, changes throughout the study, and how the relationships among CAF aspects evolve. This section discusses the results in light of the research questions and relevant literature.

5.1 The development of CAF aspects

The first research question sought to identify potential differences in CAF-based oral performance over the study period. The analysis highlighted that accuracy (percentage of error-free clauses) and speed fluency (syllables per minute) showed statistically significant improvements from phase 1 to phase 2. This suggests that by the semester's end, EFL learners spoke faster and with greater accuracy. These results align with previous L2 studies examining oral speech development during study abroad. Such studies have consistently observed increased speech rates at the end of the study abroad periods, suggesting that language learners can achieve faster speech after such exposure (D'Amico 2012; De Jong et al. 2012; Di Silvio et al. 2016; Du 2013; Llanes and Muñoz 2009; Mora and Valls-Ferrer 2012; Segalowitz and Freed 2004). This speed development was evident in both short stays, lasting three to four weeks (Llanes and Muñoz 2009), and during longer semesters (Segalowitz and Freed 2004). These patterns suggest that comparable trends in EFL oral development might be observed over a semester of online study.

While many past L2 speech development studies focused predominantly on fluency (e.g., De Jong et al. 2012; Di Silvio et al. 2016; Hanzawa 2021; Llanes and Muñoz 2009; Mora and Valls-Ferrer 2012), this study broadened its scope to other CAF dimensions and found a significant improvement in accuracy, measured by the percentage of error-free clauses. This finding is in line with the assumption that suggests various learning contexts, such as study abroad and immersion programs, tend to direct learners' attention towards grammatical forms, even when emphasizing communicative aspects (Freed et al. 2004). Changes in L2 oral production can be theoretically understood in terms of automatization. As language develops, certain speech processes demand less attention (DeKeyser 1997, 2001; Robinson 1997; Segalowitz 2010; Tavakoli 2019). The increase in accuracy and speech rate can be ascribed to this automatization, freeing up attentional resources for monitoring, thereby ensuring L2 speech accuracy (Albarqi and Tavakoli 2023; Kormos 2000). The accuracy improvement signals development in underlying speech processes, especially within the Formulator and Monitor functions (Albarqi and Tavakoli 2023; Kormos 2000, 2006; Wang 2014). This could imply that speech processing became more efficient throughout the study. Some L2 researchers argue that accuracy development often manifests in learners with already higher proficiency (Leonard and Shea 2017; Polat and Kim 2014; Rees and Klapper 2008). This study, however, did not track the pattern of changes in EFL learners with various proficiency levels. Hence, future studies might delve into EFL

oral development across different proficiency levels during extended online instruction.

5.2 The relationships among CAF aspects over time

The second research question sought to examine how the relationships among CAF measures evolve during technology-mediated TBLT instruction. The findings indicated that accuracy, lexical diversity, and syntactic complexity developed independently. That is, their development was not closely linked, suggesting that changes in one of these aspects do not necessarily lead to corresponding changes or improvements in the other aspects. McManus et al. (2021) tracked the changes in the relationships among CAF aspects for 21 months. Their findings indicated that the relationship between accuracy and complexity only emerged in the last stage of data collection. This suggests that the relationships between accuracy and complexity may need extended time to develop. Likewise, Leonard and Shea (2017) identified positive links between fluency, lexical complexity, and accuracy in the pre-test stage. However, after three months of studying abroad, they did not find notable correlations among CAF measures. They proposed that connections between these aspects might emerge over time with extended language exposure.

Fluency features, on the other hand, were intertwined with other CAF aspects across the two phases. In the first phase, strong negative correlations were noted between lexical diversity and silent pauses, repair, and syllables per minute. This indicates that ELF learners who produced a broader range of lexical elements tended to speak more slowly and exhibited fewer silent pauses and instances of self-repair. Furthermore, a strong positive correlation was observed between syntactic complexity and both syllables per minute and silent pauses. This means that utterances with higher syntactic complexity might be produced more quickly and be accompanied by silent pauses. Certain relationships remained consistent over the two phases. Specifically, accuracy maintained a consistent correlation with a reduced number of filled pauses, indicating that learners with a higher accuracy rate made fewer filled pauses in their speech. In the second phase, accuracy had also a negative correlation with repair but a positive one with syllables per minute, suggesting that more accurate speech in this phase was delivered more swiftly and included fewer instances of repairs and filled pauses. This result aligns with prior research, suggesting that an increase in accuracy rate combined with a decrease in filled pauses are hallmarks of L2 proficiency development (Albarqi and Tavakoli 2023). However, Albarqi and Tavakoli (2023) did not explore the

development of this relationship over time, leaving it unclear whether the relationship pertains to the efficiency of L2 self-monitoring or other personal or contextual influences. In the current study, the accuracy rate significantly improved throughout the study. The consistent relationship between accuracy and filled pauses was complemented in the second phase by a negative correlation with repair and a positive one with syllables per minute. These relationships may indicate development in the Formulator and self-monitoring functioning, aligning with the studies of Albarqi and Tavakoli (2023), Kormos (2000, 2006), and Wang (2014).

Over time, a persistent negative correlation was also observed between lexical diversity and repair, with a wider range of lexical elements correlating with fewer repairs. Similarly, McManus et al. (2021) identified significant and stable connections between fluency and lexis. These results indicate that lexical diversity may be closely linked with specific aspects of fluency. Moreover, the results indicated a consistent positive association between repair and breakdown features (either filled or silent pauses) across the two phases. This association aligns with self-monitoring research (see Albarqi and Tavakoli 2023; Kormos 2000, 2006; Levelt 1989), which posits that when speakers detect an error in their speech, they pause – using either filled or silent pauses – to plan their subsequent self-correction. This pattern of self-repair and pausing stems from L2 learners' limited L2 knowledge as they need to monitor different aspects of their speech (Kormos 2006). In the EFL context, self-monitoring is crucial, as it not only illuminates the gaps in learners' understanding of the target language but also prompts deeper learning processes (Kormos 2006).

The last long-term relationship in the findings was between syllables per minute and silent pauses. The finding regarding the relationship between syllables per minute and silent pauses is somewhat ambiguous as silent pauses are supposed to decrease with proficiency development (Hanzawa 2021; Tavakoli 2011), and syllables per minute to increase (Llanes and Muñoz 2009; Mora and Valls-Ferrer 2012; Segalowitz and Freed 2004). One possible explanation is that the current study did not examine mid-clause and end-clause pauses, which may yield important data. Previous research has shown that mid-clause pauses are associated with lower proficiency, while end-clause pauses are associated with higher proficiency levels (Tavakoli 2011). Future research should consider investigating these aspects of EFL fluency along with the development of speech rate within a technology-mediated TBLT context.

In summary, the results revealed that after two months of technology-mediated TBLT instruction, EFL learners exhibited significant improvements in the accuracy and speech rate of their oral production. These improvements suggest development in the underlying speech processes. While certain CAF dimensions, specifically

accuracy, syntactic complexity, and lexical diversity, developed independently, fluency features were intertwined with other CAF dimensions. Additionally, persistent relationships were observed among some CAF aspects: consistent links between accuracy and fewer filled pauses; a broader lexical range associated with fewer repairs; and a correlation between syllables per minute and silent pauses. These long-term relationships offer insights into the functioning of the underlying speech processes.

6 Conclusions

As our understanding of the impact of technology-mediated TBLT instruction on L2 learning and development suffers from a dearth of studies, particularly in the context of the study, the present research makes a significant contribution to the existing literature in several ways. Firstly, it fills a gap in the literature by examining the impact of online instruction within the framework of TBLT. The findings indicated that notable variations were observed in EFL learners' oral performance as evaluated by CAF. Specifically, it was found that their speech indicated a significant improvement in terms of both speed and accuracy by the end of the two months. This suggests that there were notable developments in their ability to articulate faster and more accurately. This study also sheds light on the development of the relationships among CAF dimensions over time. The correlation analyses revealed that accuracy, lexical diversity, and syntactic complexity developed independently. It has been assumed that the lack of negative correlations among CAF dimensions suggests the absence of a trade-off among these elements (Vercellotti 2017). That is, L2 learners do not emphasize one aspect at the cost of others; instead, these elements develop concurrently (Vercellotti 2017).

Furthermore, the study contributes to the field by providing empirical evidence on the effectiveness of technology-mediated TBLT instruction in an EFL context, demonstrating its potential to create an authentic setting in an EFL classroom and improve language learning outcomes. As the main objective of language instruction is to draw learners' attention to both form and meaning (Skehan 1998), technology-mediated TBLT can encourage L2 learners to focus on both form and meaning. However, future research needs to study L2 oral production over a longer period with a larger group of L2 learners. It also needs to frequently document the changes in oral performance to highlight patterns of development in L2 production.

The findings of this study should, nonetheless, be interpreted with caution due to the small sample size which may limit the generalizability of the findings to a larger population of L2 learners.

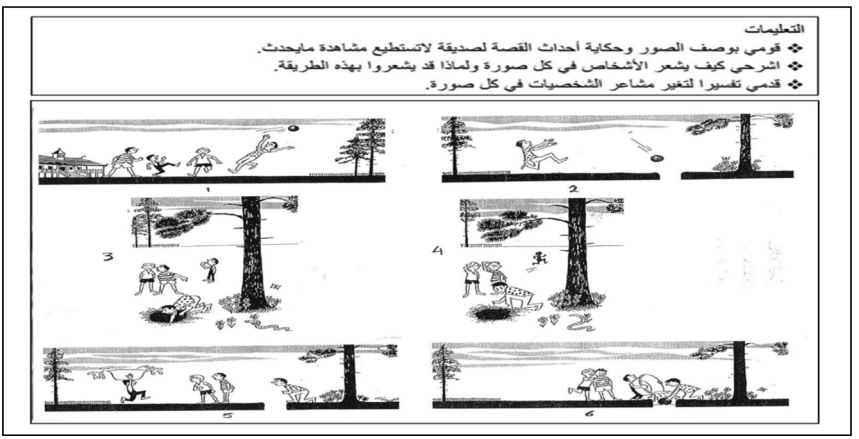
Appendix A: Tasks

a. Personal information task

Please answer the following questions. You have to complete the task.

- What do you usually do over the weekend? And how do you feel?
- What do you ideally like to do over the weekend? And how would you feel?

b. Picture prompt (Football)



c. Picture prompt (Museum)

Appendix B: Example of technology-mediated speaking task

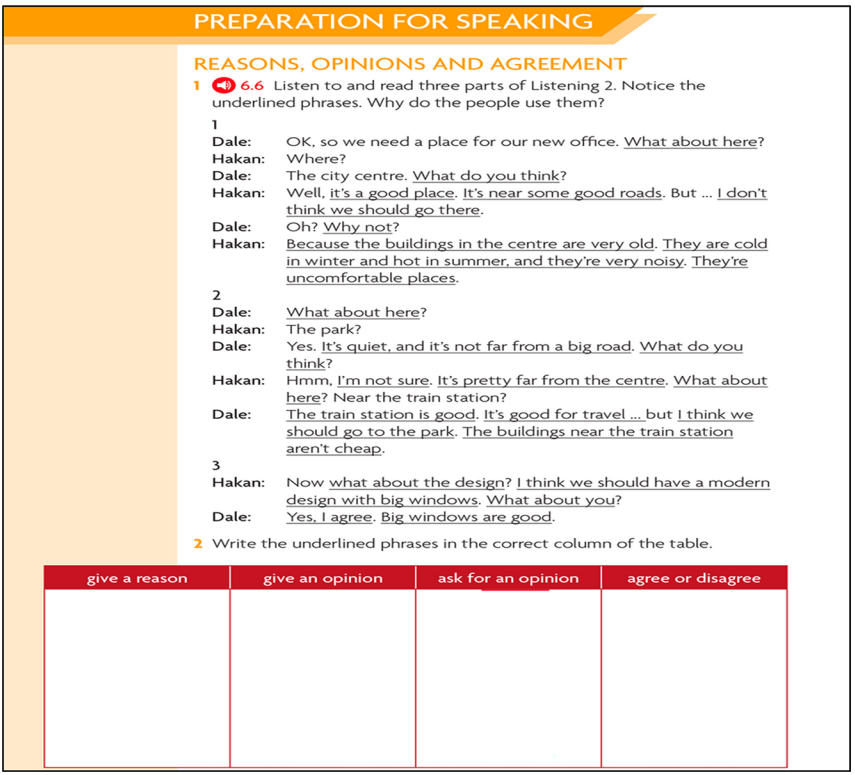
a. Task Phases aligned with technology mediated instruction

Stages	Instructions	Technological tools
Pre-task	Introducing different types of cafes	
Step 1	The instructor presents pictures of cafes constructed from various materials and featuring different designs. Then, the instructor provides questions for students to discuss in groups of five, such as matching the café's description to the corresponding picture. In this collaborative setting, students work together to solve the task and then share their answers as a group.	Class collaborate ultra eBook (presentation plus) Breakout groups Chat box Microphone Interactive white board
Step 2	The instructor introduces a set of adjectives to describe cafes, including terms such as cheap, expensive, modern, noisy, and quiet. Following this introduction, the instructor prompts students to express their preferences by choosing which type of cafe they would like to visit and providing reasons for their choices. Students are encouraged to support their opinions with reasons during the discussion.	Class collaborate ultra eBook (presentation plus) Chatbox Microphone Breakout groups WhatsApp groups Interactive white board
Step 3	The instructor plays recordings of people discussing the need for a new office (see Appendix B). Subsequently, the instructor asks students to read the transcript and identify the underlined phrases, focusing on expressions related to providing reasons, giving opinions, asking for opinions, and agreeing or disagreeing. Following this, the instructor guides students to collaborate in groups, tasking them with categorizing the identified phrases into the appropriate columns of a table found in their textbooks. Students voluntarily share their answers, seek clarification on certain phrases, and the instructor provides necessary explanations and support.	Class collaborate ultra Audio recording Chatbox eBook (presentation Plus) Microphone Breakout groups WhatsApp groups
Step 4	Moving forward, the instructor introduces a map of Green Town, featuring four potential locations for a café (near the busy main road, in the town, on the beach, on the Green Town Island) (refer to Appendix C). Students are then instructed to work in groups, selecting one of the four locations for their café and creating an outline for their discussion. The outline should highlight both positive and	

(continued)

Stages	Instructions	Technological tools
	negative aspects, and students are required to post their outlines on the Padlet website. Working collaboratively, students share their outlines on Padlet. The instructor reviews their posts on Padlet and provides constructive feedback.	
On-task	Students discuss ideas for a new café's location	
Step 5	**Task:** The instructor organizes breakout groups and assigns students the task to collaborate on. Throughout this process, she visits the groups to offer clarification and support. In the event of technical issues during the breakout groups, students also have the option to work together in WhatsApp groups for seamless communication.	Class collaborate ultra Presentation plus slides Breakout groups WhatsApp groups Microphone Chatbox Padlet Interactive white board
Step 6	**Planning:** This stage includes planning and repetition until they reach an agreement. Each group discusses their ideas for a café. They focus on completing the task by giving their opinions and providing convincing reasons. Additionally, there is a focus on maintaining a polite tone when expressing disagreements with others.	Class collaborate ultra Breakout groups WhatsApp groups Padlet Microphone Chat box
Step 7	**Report:** During this stage, the instructor ends the breakout groups and students return to the main room. They present their discussion to the entire class.	Class collaborate ultra Microphone Slides Screen sharing
Post-task	Feedback and evaluation	
Step 8	The instructor asks students to compare their answers with a different group, requiring them to specify whether they agree or disagree. They are required to provide reasons for both agreement and disagreement. The instructor assesses task completion, politeness in disagreement, pronunciation and accuracy of the language used by each group. Final comments are provided on group discussions. Finally, the instructor requires students to discuss their ideas for a café design and present their ideas to the class.	Class collaborate ultra Chat box Microphone Interactive whiteboard

b. Pre-task activity



PREPARATION FOR SPEAKING

REASONS, OPINIONS AND AGREEMENT

1 6.6 Listen to and read three parts of Listening 2. Notice the underlined phrases. Why do the people use them?

1

Dale: OK, so we need a place for our new office. What about here?
Hakan: Where?
Dale: The city centre. What do you think?
Hakan: Well, it's a good place. It's near some good roads. But ... I don't think we should go there.
Dale: Oh? Why not?
Hakan: Because the buildings in the centre are very old. They are cold in winter and hot in summer, and they're very noisy. They're uncomfortable places.

2

Dale: What about here?
Hakan: The park?
Dale: Yes. It's quiet, and it's not far from a big road. What do you think?
Hakan: Hmm, I'm not sure. It's pretty far from the centre. What about here? Near the train station?
Dale: The train station is good. It's good for travel ... but I think we should go to the park. The buildings near the train station aren't cheap.

3

Hakan: Now what about the design? I think we should have a modern design with big windows. What about you?
Dale: Yes, I agree. Big windows are good.

2 Write the underlined phrases in the correct column of the table.

give a reason	give an opinion	ask for an opinion	agree or disagree

c. A map of Green Town

You are going to open a new café in Green Town. Green Town is a small town near the sea. There are two busy roads near the town. One road goes to the capital and the other goes to the airport. Green Town is very popular with tourists. Tourists come from the capital and from countries around the world. There are ten big hotels on the beach. Green Town Island is also popular. Many tourists go on a day trip to the island. Other tourists stay in one of the island's three small hotels. Here are four places for your café:

a near the busy main road
b in the town
c on the beach
d on Green Town Island

References

Abrams, Zsuzsanna Ittzes. 2003. The effect of synchronous and asynchronous CMC on oral performance in German. *The Modern Language Journal* 87(2). 157–167.

Albarqi, Ghadah Ahmad. 2020. Self-monitoring behaviour of L2 learners: Proficiency level, dual task paradigm and working memory capacity. Unpublished PhD dissertation: University of Reading.

Albarqi, Ghadah & Parvaneh Tavakoli. 2023. The effects of proficiency level and dual-task condition on L2 self-monitoring behavior. *Studies in Second Language Acquisition* 45(1). 212–233.

Baralt, Melissa. 2014. Task complexity and task sequencing in traditional versus online language classes. In Mellisa Baralt, Roger Gilabert & Peter Robinson (eds.), *Task sequencing and instructed second language learning*, 59–122. London & New York: Bloomsbury Academic.

Baralt, Melissa & Jose Morcillo Gomez. 2017. Task-based language teaching online: A guide forteachers. *Language, Learning & Technology* 21(3). 28–43.

Boersma, Paul & David Weenink. 2008. Praat, a system for doing phonetics by computer. *Computer Software*.

Bygate, Martin. 2016. Sources, developments and directions of task-based language teaching. *Language Learning Journal* 44(4). 381–400.

Carless, David. 2012. TBLT in EFL settings: Looking back and moving forward. In Ali Shehadeh & Christine Coombe (eds.), *Task-based language learning and teaching in EFL contexts: Research and implementation*. Amsterdam & Philadelphia: John Benjamins.

Chen, Kate Tzu-Ching. 2021. The effects of technology-mediated TBLT on enhancing the speaking abilities of university students in a collaborative EFL learning environment. *Applied Linguistics Review* 12(2). 331–352.

Chong, Sin Wang & Hayo Reinders. 2020. Technology-mediated task-based language teaching: A qualitative research synthesis. *Language, Learning & Technology* 24(3). 70–86.

Cohen, Jacob. 1988. *Statistical power analysis for the behavioural sciences*, 2nd edn. Mahwah: Lawrence Erlbaum Associates.

D'Amico, Melanie Lynn. 2012. L2 fluency and willingness to communicate: The impact of short-term study abroad versus at-home study. *US-China Foreign Language* 10. 1608–1625.

De Jong, Nel & Charles A. Perfetti. 2011. Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language and learning* 61(2). 533–568.

De Jong, Nivja H., Margarita P. Steinel, Arjen Florijn, Rob Schoonen & Jan H. Hulstijn. 2012. Linguistic skills and speaking fluency in a second language. *Applied PsychoLinguistics* 34. 893–916.

DeKeyser, Robert. 1997. Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition* 19(2). 195–221.

DeKeyser, Robert. 2001. Automaticity and automatization. In Peter Robinson (ed.), *Cognition and second language instruction*, 125–151. Oxford: Oxford University Press.

Di Silvio, Francesca, Wenhao Diao & Anne Donovan. 2016. The development of L2 fluency during study abroad: A cross-language study. *The Modern Language Journal* 100(3). 610–624.

Dimond-Bayir, Stephanie, Kimberley Russel & Chris Sowton. 2019. *Unlock 2: Listening, speaking & critical thinking*. Cambridge: Cambridge University Press.

Du, Hang. 2013. The development of Chinese fluency during study abroad in China. *The Modern Language Journal* 97. 131–143.

Ellis, Rod. 2003. *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, Rod. 2008. *The study of second language acquisition*, 2nd edn. Oxford: Oxford University Press.

Ellis, Rod. 2009. Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics* 19. 221–246.
Ellis, Rod. 2018. *Reflections on task-based language teaching*, 125. Bristol, UK: Multilingual Matters.
Ellis, Rod & Gary Patrick Barkhuizen. 2005. *Analysing learner language*. Oxford: Oxford University PressMatters.
Ellis, Rod, Peter Skehan, Shaofeng Li, Natsuko Shintani & Craig Lambert. 2019. *Task-based language teaching: Theory and practice*. Cambridge: Cambridge University Press.
Faez, Farahnaz & Parvaneh Tavakoli. 2019. *Task-based language teaching*. Alexandria, Virginia: TESOL Press.
Ferrari, Stefania. 2012. A longitudinal study of complexity, accuracy and fluency variation in second language development. In Alex Housen, Folkert Kuiken & Ineke Vedder (eds.), *Dimensions of L2 performance and proficiency: Complexity accuracy and fluency in SLA*, 277–298. John Benjamins Publishing Company.
Foster, Pauline & Parvaneh Tavakoli. 2009. Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language and learning* 59(4). 866–896.
Foster, Pauline, Alan Tonkyn & Gillian Wigglesworth. 2000. Measuring spoken language: A unit for all reasons. *Applied Linguistics* 21. 354–375.
Freed, Barbara F., Norman Segalowitz & Dan P. Dewey. 2004. Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition* 26(2). 275–301.
Garcia-Ponce, Edgar Emmanuell & Parvaneh Tavakoli. 2022. Effects of task type and language proficiency on dialogic performance and task engagement. *System* 105. 102734.
González-Lloret, Marta & Lourdes Ortega (eds.). 2014. *Technology-mediated TBLT*. Philadelphia, PA: John Benjamins.
Hampel, Regine & Mirjam Hauck. 2004. Towards an effective use of audio conferencing in distance language courses. *Language, Learning & Technology* 8(1). 66–82.
Hanzawa, Keiko. 2021. Development of second language speech fluency in foreign language classrooms: A longitudinal study. *Language Teaching Research* 28(3). 816–838.
Heaton, John Brian. 1966. *Composition through pictures*. Essex: Longman.
Hokamura, Michiyo. 2018. The dynamics of complexity, accuracy, and fluency: A longitudinal case study of Japanese learners' English writing. *JALT Journal* 40(1). 23–46.
Housen, Alex, Folkert Kuiken & Ineke Vedder (eds.). 2012. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, 32. Amsterdam: John Benjamins Publishing.
Hsu, Hsiu-Chen. 2015. The effect of task planning on L2 performance and L2 development in text-based synchronous computer-mediated communication. *Applied Linguistics* 38(3). 359–385.
Kim, You Jin & Nicole Tracy-Ventura. 2013. The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System* 41(3). 829–840.
Kormos, Judit. 2000. The timing of self-repairs in second language speech production. *Studies in Second Language Acquisition* 22(2). 145–167.
Kormos, Judit. 2006. *Speech production and second language acquisition*. New York: Routledge.
Kormos, Judit & Mariann Dénes. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System* 32(2). 145–164.
Kuiken, Folkert & Ineke Vedder. 2008. Task complexity and measures of linguistic performance in L2 writing. *International Review of Applied Linguistics in Language Teaching* 45(3). 261–284.
Kusyk, Meryl. 2017. The development of complexity, accuracy and fluency in L2 written production through informal participation in online activities. *CALICO Journal* 34(1). 75–96.

Lambert, Craig & Judit Kormos. 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.

Leonard, Karen Ruth & Christine E. Shea. 2017. L2 speaking development during study abroad: Fluency, accuracy, complexity, and underlying cognitive factors. *The Modern Language Journal* 101. 179–193.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*, vol. 1. Cambridge: MIT Press.

Li, Cha & Lawrence Jun Zhang. 2023. The development of accuracy and fluency in second language (L2) speaking related to self-efficacy through online scaffolding: A latent growth curve modeling analysis. *Journal of Psycholinguistic Research* 52(5). 1371–1395.

Llanes, Àngels & Carmen Muñoz. 2009. A short stay abroad: Does it make a difference? *System* 37. 353–365.

Long, Michael. 1996. The role of the linguistic environment in second language acquisition. In William C. Ritchie & Tej K. Bhatia (eds.), *Handbook of second language acquisition*, 413–468. New York: Academic Press.

Long, Michael. 2015. *Second language acquisition and task-based language teaching*. London & England: Willey.

McManus, Kevin, Rosamond Mitchell & Nicole Tracy-Ventura. 2021. A longitudinal study of advanced learners' linguistic development before, during, and after study abroad. *Applied Linguistics* 42(1). 136–163.

Mora, Joan C. & Margalida Valls-Ferrer. 2012. Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TesoI Quarterly* 46. 610–641.

Norris, John M. & Lourdes Ortega. 2009. Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics* 30(4). 555–578.

Pallotti, Gabriele. 2009. CAF: Defining, refining and differentiating constructs. *Applied Linguistics* 30(4). 590–601.

Park, Sujung. 2010. The influence of pretask instructions and pretask planning on focus on form during Korean EFL task-based interaction. *Language Teaching Research* 14. 9–26.

Patanasorn, Chomraj. 2010. Effects of procedural content and task repetition on accuracy and fluency in an EFL context. Unpublished PhD dissertation: Northern Arizona University.

Polat, Brittany & Youjin Kim. 2014. Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied Linguistics* 35. 184–207.

Préfontaine, Yvonne & Judit Kormos. 2016. A qualitative analysis of perceptions of fluency in second language French. *International Review of Applied Linguistics in Language Teaching* 54(2). 151–169.

Rees, Jonathan & John Klapper. 2008. Issues in the quantitative longitudinal measurement of second language progress in the study abroad context. In Lourdes Ortega & Heidi Byrnes (eds.), *The longitudinal study of advanced L2 capacities*, 89–105. New York: Routledge.

Robinson, Peter. 1997. Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition* 19(2). 223–247.

Robinson, Peter. 2001. Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics* 22(1). 27–57.

Robinson, Peter. 2007. Task complexity, theory of mind, and intentional reasoning: Effects of L2 speech production, interaction, uptake and perceptions of task complexity. *International Review of Applied Linguistics in Language Teaching* 45. 193–213.

Robinson, Peter. 2011. *Task-based language learning*, 5. New Jersey: John Wiley & Sons.

Satar, Müge & Nesrin Özdener. 2008. The effects of synchronous CMC on speaking proficiency and anxiety: Text versus voice chat. *The Modern Language Journal* 92(4). 595–613.

Segalowitz, Norman. 2010. *Cognitive bases of second language fluency*. London: Routledge.
Segalowitz, Norman & Barbara F. Freed. 2004. Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition* 26(2). 173–199.
Skehan, Peter. 1998. *A cognitive approach to language learning*. Oxford: Oxford University Press.
Skehan, Peter. 2009. Modeling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics* 30. 510–532.
Skehan, Peter. 2014. Limited attentional capacity, second language performance, and task-based pedagogy. In Peter Skehan (ed.), *Processing perspectives on task performance*, 211, 211–260. Amsterdam: John Benjamins Publishing Company.
Skehan, Peter, Pauline Foster & Sabrina Shum. 2016. Ladders and snakes in second language fluency. *International Review of Applied Linguistics in Language Teaching* 54. 97–111.
Smith, Bryan & Marta González-Lloret. 2021. Technology-meidated task-based language teaching: A research agenda. *Language Teaching* 54(4). 518–534.
Spoelman, Marianne & Marjolijn Verspoor. 2010. Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics* 31(4). 532–553.
Suzuki, Shungo & Judit Kormos. 2023. The multidimensionality of second language oral fluency: Interfacing cognitive fluency and utterance fluency. *Studies in Second Language Acquisition* 45(1). 38–64.
Swain, Merrill. 2005. The output hypothesis: Theory and research. In Eli Hinkel (ed.), *Handbook of research in second language teaching and learning*, 471–484. Mahwah & New Jersey: Lawrence Erlbaum Associates.
Tavakoli, Parvaneh. 2011. Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal* 65(1). 71–79.
Tavakoli, Parvaneh. 2019. Automaticity, fluency and second language task performance. In Edward Wen Zhisheng & Mohammad Ahmadian (eds.), *Researching L2 task performance and pedagogy: In honour of Peter Skehan*. Amsterdam: John Benjamins.
Tavakoli, Parvaneh & Pauline Foster. 2011. Task design and second language performance: The effect of narrative type on learner output. *Language Learning* 61(1). 37–72.
Tavakoli, Parvaneh & Peter Skehan. 2005. Strategic planning, task structure, and performance testing. In Rod Ellis (ed.), *Planning and task performance in a second language*, 11, 239–273. Amsterdam: John Benjamins.
Tonkyn, Alan P. 2012. Measuring and perceiving changes in oral complexity, accuracy and fluency: Examining instructed learners' short-term gains. In Alex Housen, Folkert Kuiken & Ineke Vedder (eds.), *Dimensions of L2 performance and proficiency: Complexity accuracy and fluency in SLA*, 221–244. Amsterdam: John Benjamins Publishing Company.
Van der Zwaard, Rose & Anne Bannink. 2016. Nonoccurrence of negotiation of meaning in task-based synchronous computer-mediated communication. *The Modern Language Journal* 100(3). 625–640.
Van Moere, Alistair. 2012. A psycholinguistic approach to oral language assessment. *Language Testing* 29(3). 325–344.
Vercellotti, Mary Lou. 2017. The development of complexity, accuracy, and fluency in second language performance: A longitudinal study. *Applied Linguistics* 38. 90–111.
Volle, Lisa M. 2005. Analyzing oral skills in voice email and online interviews. *Language, Learning & Technology* 9(3). 146–163.
Wang, Yuping. 2004. Supporting synchronous distance language learning with desktop videoconferencing. *Language, Learning & Technology* 8. 90–121.

Wang, Zhan. 2014. On-line time pressure manipulations. In Peter Skehan (ed.), *Processing perspectives on taskperformance*, 5, 27–61. Amsterdam: John Benjamins.
Westbrook, Carolyn, Lida Baker & Chris Sowton. 2019. *Unlock 2: Reading, writing & critical thinking*. Cambridge: Cambridge University Press.
Willis, Jane. 1996. *A framework for task-based teaching*. Harlow & England: Longman.
Willis, David & Jane Willis. 2007. *Doing task-based teaching: A practical guide to task-based teaching for ELT training courses and practising teachers*. Oxford & England: Oxford University Press.
Yu, Hanjing & Wander Lowie. 2020. Dynamic paths of complexity and accuracy in second language speech: A longitudinal case study of Chinese learners. *Applied Linguistics* 41(6). 855–877.
Ziegler, Nicole. 2016. Taking technology to task: Technology-mediated TBLT, performance, and production. *Annual Review of Applied Linguistics* 36. 136–163.

Bionote

Ghadah Albarqi
Foreign Language Department, Taif University, Taif, Saudi Arabia
gadah.g@tu.edu.sa
https://orcid.org/0000-0001-9043-5881

Dr Ghadah Albarqi, an Assistant Professor of Applied Linguistics at Taif University, holds PhD in Second Language Acquisition from the University of Reading. Dr Albarqi's research interests include second language production, assessment and technology-mediated TBLT. Her recent research includes:
Albarqi, G., & Tavakoli, P. (2023). The effects of proficiency level and dual-task condition on L2 self-monitoring behavior. *Studies in Second Language Acquisition*, *45*(1), 212–233.
Albarqi, G. (2023). Padlet as a Formative Assessment Tool in the Online Language Classroom: Action Research. In: Chong, S.W., Reinders, H. (eds) *Innovation in Learning-Oriented Language Assessment. New Language Learning and Teaching Environments.* Palgrave Macmillan, Cham.