The Language Learning Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rllj20

Complexity, accuracy, and fluency improvements through massively multiplayer online gaming: a longitudinal mixed-methods case study

Nasser Jabbari & Mark Peterson

To cite this article: Nasser Jabbari & Mark Peterson (2023): Complexity, accuracy, and fluency improvements through massively multiplayer online gaming: a longitudinal mixed-methods case study, The Language Learning Journal, DOI: 10.1080/09571736.2023.2219713

To link to this article: https://doi.org/10.1080/09571736.2023.2219713

Published online: 11 Jul 2023.

THE LANGUAGE LEARNING JOURNAL
https://doi.org/10.1080/09571736.2023.2219713

Complexity, accuracy, and fluency improvements through massively multiplayer online gaming: a longitudinal mixed-methods case study

Nasser Jabbari [a] and Mark Peterson [b]

[a]Department of Language and Linguistics, University of Essex, Colchester, UK; [b]Department of Foreign Language Studies Graduate School of Human and Environmental Studies, Kyoto University, Kyoto, Japan

ABSTRACT

Many English as a foreign language (EFL) learners worldwide spend hours playing commercial-off-the-shelf massively multiplayer online games. They engage in a multitude of different interactions with the game environment and possibly with other gamers, using English as the predominant medium of communication. This raises the question of whether playing such games improves these learners'/gamers' speaking performance over time. Drawing on usage-based theories of language learning, this longitudinal mixed-methods case study addressed this question by measuring six EFL gamers' speaking performance in terms of complexity, accuracy, and fluency (CAF) over the course of six months, and exploring their perceptions relating to the gaming environment and their language learning. The participants played in two teams: Team 1 with low-intermediate and Team 2 with high-intermediate proficiency. Each team comprised three EFL gamers and a native English speaker. Overall, the results appeared positive for the EFL gamers on both teams but in different ways. Their speaking performance showed improvements in 50% of the CAF indices. These results are discussed in light of the gamers' perceptions about the dynamics of in-game communication, the game's socio-affective environment, and the development of their speaking performance.

ARTICLE HISTORY
Received 13 February 2023
Accepted 25 May 2023

KEYWORDS
Massively multiplayer online games; speaking performance; complexity; accuracy; fluency

Introduction

Does it still seem 'surreal' to envision a future where digital games function as normalised global venues for second language (L2) learning and socialisation? This no longer seems to be the case for numerous language researchers and educators. Considering the quantity, quality, and diversity of second language (L2) use in, around, and about digital games and the transformation of these games into a global, multiplayer, and often multilingual practice, 'it has become easier to imagine digital games as authentic, consequential, and widely applicable L2 learning resources' (Reinhardt and Thorne 2016: 416). Commercial-off-the-shelf (COTS) massively multiplayer online games (MMOGs) have been integrated into many people's daily schedules around the globe (Yee 2006; Reinhardt and Thorne 2016). These games replicate a two- or three-dimensional persistent virtual world where gamers can interact – through personalised in-game characters or 'avatars'– with the

CONTACT Nasser Jabbari n.jabbari@essex.ac.uk

game environment and potentially with thousands of gamers worldwide. Gamers must complete an increasingly challenging sequence of tasks or 'quests' to progress to a higher level; and to accomplish higher-level goals, they need to collaborate with their game fellow players in small or large groups called 'guilds'. Efficient and effective multimodal communication (e.g. through text and/or voice chat channels) leads to success in MMOGs. These games have gained tremendous popularity in recent years; for example, *World of Warcraft* (*WoW*) is estimated to have about 109.34 million subscribers and 4.16 million active players per day as of 2020 (MMO Populations, n.d.). Research (e.g. Yee 2006) shows that people of different ages and gender spend millions of hours playing MMOGs, producing and sharing knowledge as well as socialising in like-minded 'communities of practice' (Lave and Wenger 1991). Myriads of gamers become immersed in such highly interactive social settings and predominantly use English, often as a L2, as a medium of communication. L2 gamers have claimed that participating in various in-game and beyond-game activities has helped them develop new L2 skills and improve their L2 proficiency (e.g. Chotipaktanasook and Reinders 2018; Peterson 2012a).

There is thus growing interest in the potential benefits of playing MMOGs for L2 development. Since their introduction into the world of entertainment, these games have attracted much scholarly interest (Hung et al. 2018); however, answers to some key questions have remained elusive. For instance, to what extent and under what circumstances (e.g. technical, social, psychological, and cognitive[1]) does MMO gameplay support L2 development? What are the underlying processes or mechanics of L2 development in such complex and multi-faceted media? Research suggests that this elusiveness partially stems from the proliferating varieties of MMOG genres and the complex nature of L2 development. L2 development is arguably a multi-faceted phenomenon that undergoes dynamic changes across time and context (Thorne and Tasker 2011; Larsen-Freeman and Cameron 2008) and this development involves an 'ongoing emergence of complexity, fluency, and accuracy in learner language' (Larsen-Freeman 2006: 590). As such, as Verspoor and Behrens (2011: 25) have argued, '[t]here is not one single theory that deals with all aspects of what language is, how it is organized, how it is processed, how it is used, how it changes, how it is acquired and how it is learned as a second language'.

Although still in its infancy, current research on COTS MMOGs (see reviews by Peterson et al. 2022; Jabbari and Eslami 2019; Li 2020) has suggested that using L2 English during gameplay and beyond the virtual worlds of the games (e.g. watching walkthroughs, reading game reviews, and discussing in online game fora) can provide useful opportunities for L2 development. These findings support the notion of 'naturalistic CALL,' which theorises L2 learning as a peripheral outcome of 'students' pursuit of some leisure interest through a second or foreign language in digital environments in informal learning contexts, rather than for the explicit purpose of learning the language' (Chik 2013: 835–836). Except for a few studies (e.g. Newgarden, Zheng, and Liu 2015; Zheng, Newgarden, and Young 2012), the existing body of research has focused predominantly on text-based communication. As many gamers now have audio or audio-visual interactions while gaming, it is important to investigate if, how, and to what extent, their L2 speaking performance improves over time. However, there is a lack of longitudinal research that has systematically examined the trends of improvements in L2 gamers' speaking performance in terms of complexity (lexical and syntactic), accuracy, and fluency (CAF) as important constructs of language development (Larsen-Freeman 2009; Norris and Ortega 2009). The research reported here seeks to address this gap by using 14 CAF measures to investigate changes in the speaking performance of two different groups of English-as-a-foreign language (EFL) gamers (with lower and higher L2 proficiency, respectively) who played *WoW* with a native English speaker (NES) over a period of six months.

Within this research, L2 development is viewed through 'usage-based' or 'emergentist' theories of language learning (Behrens 2009; Tomasello 2003). From this perspective, language development is an iterative process and evolves as the result of extensive use in meaningful interactions with the environment. As such, the frequency of language use and the emerging linguistic patterns are salient indicators of language development (Ellis 2002). Given this specific context, this study endeavours to answer the following research questions:

Q1: How do measures of speaking performance (lexical and syntactic complexity, morphosyntactic accuracy, and fluency) of EFL gamers change over six months of sustained collaborative playing of *WoW*?
Q2: Are there differences in the change patterns of gamers with lower vs. higher EFL proficiency?

Literature review

Network-based computer-mediated communication settings are considered potentially beneficial environments for L2 learners to develop and reinforce L2 skills (Chapelle 2008) and this view can be extended to online digital games, particularly MMOGs as 'complex semiotic ecologies' (Thorne, Fischer, and Lu 2012) that afford multiple opportunities for L2 learning and socialisation (Peterson and Jabbari 2022). Since the birth of *Neverwinter Nights* as the first fully graphical multiplayer role-playing game in 1991, SLA scholars have sought to explore the potential of these still-evolving online social settings (Dixon and Christison 2021; Peterson 2016). Researchers (e.g. Reinhardt and Thorne 2020) argue that using an L2 meaningfully and purposefully in order to complete various game-mediated and game-related tasks can improve L2 proficiency. They argue that MMOGs are semiotically rich and socially interactive settings that incorporate substantial opportunities 'for language socialisation and for acquisition of skills related to just-in-time linguistic tools and services' (Godwin-Jones 2014: 12). Most recent reviews (e.g. Jabbari and Eslami 2019; Peterson 2021) and meta-analyses (e.g. Dixon, Dixon, and Jordan 2022) have found that participating in gameplay – within and beyond the game environment – can promote autonomous language learning practices (e.g. Chik 2014; Ibrahim 2018; Li, Peterson, and Wang, 2022), improve L2 vocabulary (e.g. Sundqvist 2019; Sundqvist and Wikström 2015), develop L2 communicative competence (e.g. Peterson 2012a; Rama et al. 2012), and enhance cross-cultural awareness (e.g. Rama et al. 2012; Thorne 2008). Research (e.g. Horowitz 2019; Li, Peterson, and Wang 2022) has also shown that MMOGs provide socially and emotionally supportive communication environments that promote positive affect and motivation, reduce L2 anxiety, boost learners' self-confidence, and enhance willingness to communicate. Some researchers (e.g. Peterson 2012a; Reinders and Wattana 2015b) have argued that providing L2 gamers with the opportunity to remain anonymous (through avatar-embodied interactions) is one of MMOGs' most significant affordances for mitigating L2 anxiety and thereby encouraging L2 gamers' risk-taking endeavours in using the L2.

A small number of studies have addressed, albeit briefly, aspects of complexity, accuracy, and fluency in L2 gamers' spoken and written performance. For example, Reinders and Wattana (2011) conducted research with 16 undergraduate students who were instructed to play *Ragnarok Online* collaboratively in three 90-minute game sessions and use L2 English to communicate during gameplay. The researchers modified the game by creating three new quests relevant to the content and objectives of the lessons the participants had studied in an English course. They found that the gameplay improved the quantity of L2 interactions (in both text- and voice-based chat channels) but did not advance the accuracy and complexity of the students' L2 utterances. They reported that the participants who communicated through text paid more attention to grammatical accuracy than those who communicated through voice.

Rankin et al. (2009) studied the impact of in-game social interactions between native and non-native English speakers (NNESs) on L2 vocabulary acquisition. The NNESs, advanced English-as-a-second language (ESL) students, were randomly assigned to three groups. They either attended 3 hours of class instruction, played *Ever Quest II* independently for 4 hours, or played the game with NESs. Regardless of treatment, 82% of the students who played the game showed increased vocabulary knowledge. The students in traditional classroom instruction scored the highest for appropriately using vocabulary in sentences. Those who played the game collaboratively with the NESs scored the highest in recognising the correct word in gameplay scenarios.

Given the scarcity of longitudinal CAF research in the context of MMOGs and the primary focus of previous research on text-based interactions, this study sets out to explore whether EFL gamers' speaking performance improved over six months based on 14 CAF indices.

Methodology

The research reported here comprises a longitudinal mixed-methods single-case study. According to Yin (2018: 49), a case study is 'an empirical method that investigates a contemporary phenomenon (the 'case') in depth and within its real-world context, especially when the boundaries between the phenomenon and context may not be clearly evident'. The phenomenon (the 'case') under investigation is EFL speaking development in the context of *WoW* gaming. As a longitudinal study, this research is designed to systematically map the developmental trends of complexity, morphosyntactic accuracy, and fluency in six EFL gamers' speaking performance over six months. We collected data from two teams of EFL gamers (with low and high intermediate L2 proficiency, respectively) using a systematic data collection procedure to allow comparisons across different data samples within and between participants. This study also shares some characteristics of a single case time-series design. As Menard (2008: 579) explained, '[a] time series is a set of repeated measurements of the same variable taken on the same unit of analysis (e.g. an individual, city, nation; more generally, a subject or a case) for two or more points in time'. This research entails repeated measurements of CAF indices across all samples of data generated during the gameplay: see Table 1. We implemented the intervention (i.e. collaborative L2 gaming) at the beginning of the data-collection period without establishing any baseline. Since linguistic development is unlikely to be discernible over a short period (Ortega 2003), we considered the first and the second data points (within the first two months) as the baseline. We expected to observe improving trends in the CAF indices over time.

Study setting: WoW

WoW is a recreational massively multiplayer online role-playing game (MMORPG) launched in 2004 by Blizzard Entertainment.[2] Like any other MMORPG, thousands of *WoW* players, located in different

Table 1. CAF indices used in the current research.

Complexity	**Lexical complexity:** • Lexical density • Lexical sophistication • Lexical variation **Syntactic complexity:** • The average number of subordinate clauses per AS-unit
Accuracy	• Percentage of erroneous clauses • Percentage of erroneous AS-units
Fluency	**Breakdown (dys)fluency:** • Average number of silent pauses of 0.5 s or more (between and within AS-units) per AS-unit • Average number of filled pauses including those of less than 0.5 s (between and within AS-units) per AS-unit • Silent pause ratio: i.e. total silent-pause duration divided by total speaking-time duration. Speaking-time duration refers to the time taken to produce speech including all hesitation phenomena (e.g. silent and filled pauses, false starts, restarts, repetitions, etc.). • Filled pause ratio: i.e. total filled-pause duration divided by total speaking-time duration. **Repair (dys)fluency:** • Number of repetitions, restarts, false starts, and repairs per minute: i.e. number of repetitions, restarts, false starts, and repairs divided by the total speaking-time duration × 60. **Speed fluency:** • Phonation time ratio: i.e. phonation time (total speaking-time duration minus the total duration of silent and filled pauses) divided by total speaking-time duration. • Articulation rate per minute: i.e. total number of syllables divided by phonation time × 60. • Speech rate per minute: i.e. total number of syllables divided by total speaking-time duration × 60.

parts of the world, interact, collaborate and compete simultaneously. *WoW* contains fantasy and science fiction elements and provides a dynamic virtual world with high-quality 3D graphics. Before starting the game, *WoW* players select the realm (or server) in which they would like to play the game. *WoW* provides different types of realms: player versus environment (PvE), role-play (RP), and role-play player versus player (RP-PvP). In this study, the participants played in the PvE realm. This realm's primary focus is defeating game-controlled monsters and completing in-game missions (or quests) alone or in collaboration with other players.

After selecting the realm, players pick and customise their in-game avatar, i.e. a three-dimensional character that mediates a player's interactions with the game world. *WoW* players choose their avatars' race and class and customise their gender and appearance. *WoW* is based on constant warfare between two opposing factions: the Alliance and the Horde. As a team of four, the participants preferred to play in the Alliance faction, which is a calm, peaceful and civilised faction compared to the Horde.

Much of the gameplay in *WoW* involves completing different types of quests assigned by game-controlled non-playing characters (NPCs). Quests cover various activities, e.g. killing computer-controlled monsters called mobs, gathering resources, and finding and delivering items to NPCs. Gameplayers can finish quests alone; however, as the game progresses, the quests grow too challenging to complete independently. Tough challenges are commonplace in 'dungeons', hostile closed-off areas in caves and castles where players face the most formidable enemies. Getting through a dungeon requires a well-trained and well-balanced team of players with appropriate levels for the challenge. A typical dungeon (made available around level 15) allows a group of five characters to enter. In this study, the participants formed two teams in the Alliance faction, each consisting of four players (one NES and three NNESs). They entered dungeons and completed the quests, using an external audio channel, TeamSpeak3®, to communicate during the gameplay.

Participants

The participants were six EFL gamers based in Iran and an NES based in Texas, USA. We decided to focus on this small number of participants in order to allow for detailed and intensive analyses of the data, which, in turn, would provide a concrete illustration of the phenomenon under investigation. The first author started the project by posting invitation-to-participate letters on his Facebook page. These letters consisted of short surveys to help with an initial screening. The researchers received responses from three NESs and 24 EFL gamers, who were then contacted and provided with more information about the project. At this point, two NESs and 12 EFL gamers opted out as they could not manage the time difference between the two countries involved. The remaining NES, a 30-year-old expert *WoW* player from Texas, signed the consent form and agreed to join the project. The researchers assessed the remaining EFL gamers' access to minimum system requirements (for playing the game on a European server) and administered an English placement test covering CEFR levels from A1 to C1. Three candidates were screened out for failing to meet the technical requirements, and the remaining three were eliminated from the current study due to their advanced English language proficiency. The remaining six signed the consent forms and agreed to play the game for at least 30 h. They were male undergraduate students (23–25 years old) who had been playing other MMOGs in L2 English on European servers. Three whose English was judged at CEFR A2/B1 level comprised the pre-intermediate team (T1) and three whose English was judged at B2 level comprised the upper-intermediate team (T2). None of the EFL gamers selected had used voice chat to communicate during their gameplay. The NES participated in both teams. To maintain participants' anonymity, the EFL gamers' initials (F, MH, and MM in Team 1; B, E, and M in Team 2) and the pseudonym 'Nate' for the NES have been used in reporting the research.

Data collection

The data for this study comprises five equally distributed samples (100 minutes each) of audio-recorded in-game conversations. These data were taken from a larger corpus of 59.96 hours of naturally-occurring in-game conversations; other data were derived from the EFL gamers' reflection journals written after each gaming session, and their post-study semi-structured interview transcripts.

The participants played the game over a period of six consecutive months; T1 completed 14 gaming sessions giving a total of 29.58 hours of game play, while T2 completed 15 sessions or 30.38 hours. The participants were instructed to all start the game simultaneously and play as a team. We did not provide any other instructions in order to ensure an authentic game-playing experience. A proprietary voice-over-Internet Protocol (VoIP) software, TeamSpeak3®, was installed for the participants to engage in real-time genuine conversations while gaming, allowing them to talk over each other's voices without having to take turns. To eliminate the impact of the researcher's presence on the participants' in-game social and linguistic behaviour, we asked Nate to audio-record all sessions.

We examined the five data samples for each player using 14 CAF measures. As noted above, the samples were produced at five equally distributed time points (hereafter t1, t2, t3, t4, and t5). As the audio files featured pauses between turns and encompassed utterances of more than one participant, we opted for long (100 min) data samples to ensure the adequacy of the samples. Thus, the data used for this research consists of 500 minutes (100 minutes × 5 data points) of audio-recorded conversation for each EFL gamer; that is 1500 minutes (25 hours) of recorded data for each team of three, and 3000 min (50 hours) of audio-recorded spoken data in total. The respective data samples consisted of 7464 words for T1 and 17834 words for T2; that is, an average of 2488 words for each EFL gamer in T1 and 5945 words for each EFL gamer in T2.

We judged participants' gaming experience as pretty uniform across the five time points. They were involved in similar actions (e.g. collaborating, competing, and socialising), tasks (e.g. collecting, killing, delivering, and trading), interactions (with other gamers and the game environment), and language use. Similarly, their gameplay style did not vary much across the five time points. They played mainly in the PvE realm, defeating monsters and completing quests in collaboration with their teammates. They also played in the PvP realm, competing against other similarly capable gamer teams in battlegrounds. At the end of the project, the participants in T1 and T2 had reached levels 40 and 43 respectively. We made the assumption that the data elicited at the five time points would justify comparisons between T1 and T2 in terms of CAF indices.

Data analyses

The in-game conversations were transcribed, coded, and scored based on the 14 CAF measures (see Table 1). The transcripts were edited to exclude extraneous words involved in false starts (e.g. *well* and *you know*), reformulations, and functionless repetitions. Further, one-word minor utterances (e.g. *yes*, *no*, *okay*, *uh-huh*, *right*, *alright*, *cool*, *well*, *hello*, and *thanks*) and echoic responses were also excluded from the transcripts. As Foster et al. (2000: 370) have indicated, highly interactional spoken data such as these recordings can 'yield a high proportion of minimal units (e.g. one-word minor utterances and echoic responses) whose inclusion in an analysis could distort the perception of the nature of the performance.' The edited versions of the transcripts were used for analysis. The basic unit of analysis was the Analysis of Speech unit (AS-unit) which has been 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, Tonkyn, and Wigglesworth 2000: 365).

To ensure the reliability of the researchers' segmentation of the transcribed audio-recorded utterances into AS-units and subordinate clauses and identification of the morphosyntactic errors, 20% of the data were segmented, coded, and scored by an independent expert. Likewise, a second coder coded the entire data to ensure the reliability (Merriam and Tisdell 2016) of the analyses and

interpretation of the interviews and reflection journals. The inter-coder reliability, using Cohen's Kappa (Cohen 1960), was calculated for the interviews and reflection journals ($\kappa = 0.90$), AS-units ($\kappa = 0.89$), subordinate clauses ($\kappa = 0.88$) and morphosyntactic errors ($\kappa = 0.94$). Cases of disagreement were discussed and resolved.

Measures of lexical complexity

Lu (2012) conceptualised lexical complexity or 'lexical richness' (Read 2000) as a multifaceted characteristic of language use that consists of lexical density, sophistication, and variation, as well as the number of errors in vocabulary use.

Lexical density, initially coined by Ure (1971), refers to 'the ratio of the number of lexical – as opposed to grammatical – words to the total number of words in a text' (Lu 2012: 191). Although lexical density is typically a measure of the linguistic complexity of written texts (Halliday 1985), it 'plays more of a role in the analysis of spoken texts' (Read 2000: 203). *Analyze My Writing* (n.d.), a free online text content and readability analyser, was used to measure lexical density.

Lexical sophistication is 'a measure of the proportion of relatively unusual or advanced words in the learner's text' (Read 2000: 203). A word frequency measure based on frequency bands (Laufer and Nation 1995; Morris and Cobb 2004) was used to measure the lexical sophistication of the EFL gamers' speaking performance. Word frequency scores were obtained by rank-ordering all words in a single master frequency list based on the British National Corpus and the Corpus of Contemporary American English (BNC-COCA 1-25k) lists.[3] Then, the percentage of the spoken data within each frequency band (i.e. 1–25k) was calculated. The freeware VocabProfilers available from the Compleat Lexical Tutor v.8.3 programme (Cobb 2020) was utilised to obtain these results.

Lexical variation (Lu 2012) is also labelled as 'lexical diversity' (Malvern et al. 2004) and 'lexical variety' (Malvern and Richards 2002). It is defined as 'the range of a learner's vocabulary as displayed in his or her language use' (Lu 2012: 192) or 'the number of different words in a sample of speech or writing of a set length' (Malvern et al. 2004: 3). This study used a new transformation of type-token ratio (TTR), i.e. the D measure (Malvern and Richards 2002; McKee, Malvern, and Richards 2000). High values of D indicate high levels of lexical variation. The D measure was used in preference to other measures of lexical variation[4] for two primary reasons. First, it 'provides a robust measure of lexical diversity which is not a function of sample size in the way raw TTR and its simple transformations are' (Malvern et al. 2004: 60). Second, its validity has been established by data samples from adult ESL learners. The D measure was calculated using the VOCD command in the Computerised Language Analysis (CLAN) programme (MacWhinney 2000).[5]

Measure of syntactic complexity

Kyle (2016: 8) has explained that syntactic complexity is 'the formal characteristics of syntax (e.g. the amount of subordination), which has been described as absolute complexity (Bulté and Housen 2012).' In this research, the average number of subordinate clauses per AS-unit was calculated to measure syntactic complexity. This measure was applied because it is calculated based on the AS-unit, which is primarily a syntactic, rather than semantic or intonational, unit of analysis. This measure can reveal the extent to which participants are able to combine smaller blocks of language (i.e. subordinate clauses) to produce more complex and extended syntactic turns (Foster et al. 2000).

Measures of morphosyntactic accuracy

The morphosyntactic accuracy of participants' speaking performance was assessed by calculating the percentages of erroneous clauses and AS-units. An error in this study constituted any non-target production of English syntax and morphology (e.g. word order, subject-verb agreement, article-noun agreement, adverb/adjective substitution). The error analysis here examined only 'deviations in correctness' and ignored 'deviations in appropriateness' (Ellis 1994: 52). Thus, errors in semantics and lexicon (meaning and vocabulary), phonology (pronunciation), and discourse (style), as well as self-corrected errors were discarded.

Measures of fluency

Skehan and Foster (1999: 96) defined fluency as 'the capacity to use language in real time, to emphasise meanings, possibly drawing on more lexicalised systems'. Like complexity and accuracy, fluency is a multidimensional construct (Housen and Kuiken 2009). Therefore, different measures are required to capture distinct aspects of this construct. The measures used in this study are based on three sub-dimensions of utterance fluency (Segalowitz 2010) proposed by Skehan (2003, 2009): *breakdown (dys)fluency*, which acknowledges hesitation phenomena (i.e. silent and filled pauses); *repair (dys)fluency*, which considers the occurrence of repetitions, reformulations, false starts, and replacements; and *speed fluency*, which incorporates temporal characteristics of speech such as syllables per minute (see Table 1). The three most extended turns (each including two or more AS-units) were extracted from each sample of data to assess speed fluency, i.e. 15 turns (3 turns × 5 samples of data) for each EFL gamer and 45 turns (15 turns × 3 EFL gamers) for each team (see Table 2). PRAAT software (Boersma and Weenink 2013) was used to measure the temporal aspects of fluency, i.e. speaking-time duration, lengths of silent and filled pauses, phonation time, and articulation rate.

Results

Due to the small sample size (i.e. three EFL gamers in each team), we could not run inferential statistics (Pearson's correlation or Spearman's rho) to test the correlations between the CAF indices and time spent playing the game across five data points. Therefore, we decided to report the results by visualising the trends in the data using scatter plots.

Complexity

Lexical density

As the trendlines in Figure 1 show, the average values of lexical density tended to increase over time in T1 but remained almost flat in T2. In Excerpts 1 and 2, participants MM and MH from T1 shed some light on this in their post-study interviews where they refer to their quick, telegraphic, and as such lexically dense dialogic interactions during the gameplay, which contributed to the team's success.

Excerpt 1 (MM's interview 2:54)

As you've noticed, we were more silent on some occasions when we were under pressure and had to make decisions in milliseconds. We were using only the keywords we needed quickly.

Excerpt 2 (MH's interview 17:10)

I had to say something as quickly as possible to get some help or advice. Sometimes, I could use one or two words [to communicate with others]. For example, when we were capturing the flag or when a boss was approaching, I had no time for a complete sentence … I used two or three words to let others know.

Lexical sophistication

The results for lexical sophistication, shown in Figure 2, suggest that, across the five time points, the EFL gamers in both groups were relying mostly on the first 1000 common words (1k) to

Table 2. The total and average sample sizes (in seconds) for speed fluency assessments.

Team 1	**F**	**MH**	**MM**
Total sample size (extracted from 5 time points)	251.15	100.59	260.43
Average sample size for a time point	50.23	20.11	52.08
Team 2	**B**	**E**	**M**
Total sample size (extracted from 5 time points)	229.51	287.78	176.82
Average sample size for a time point	45.90	57.55	35.36

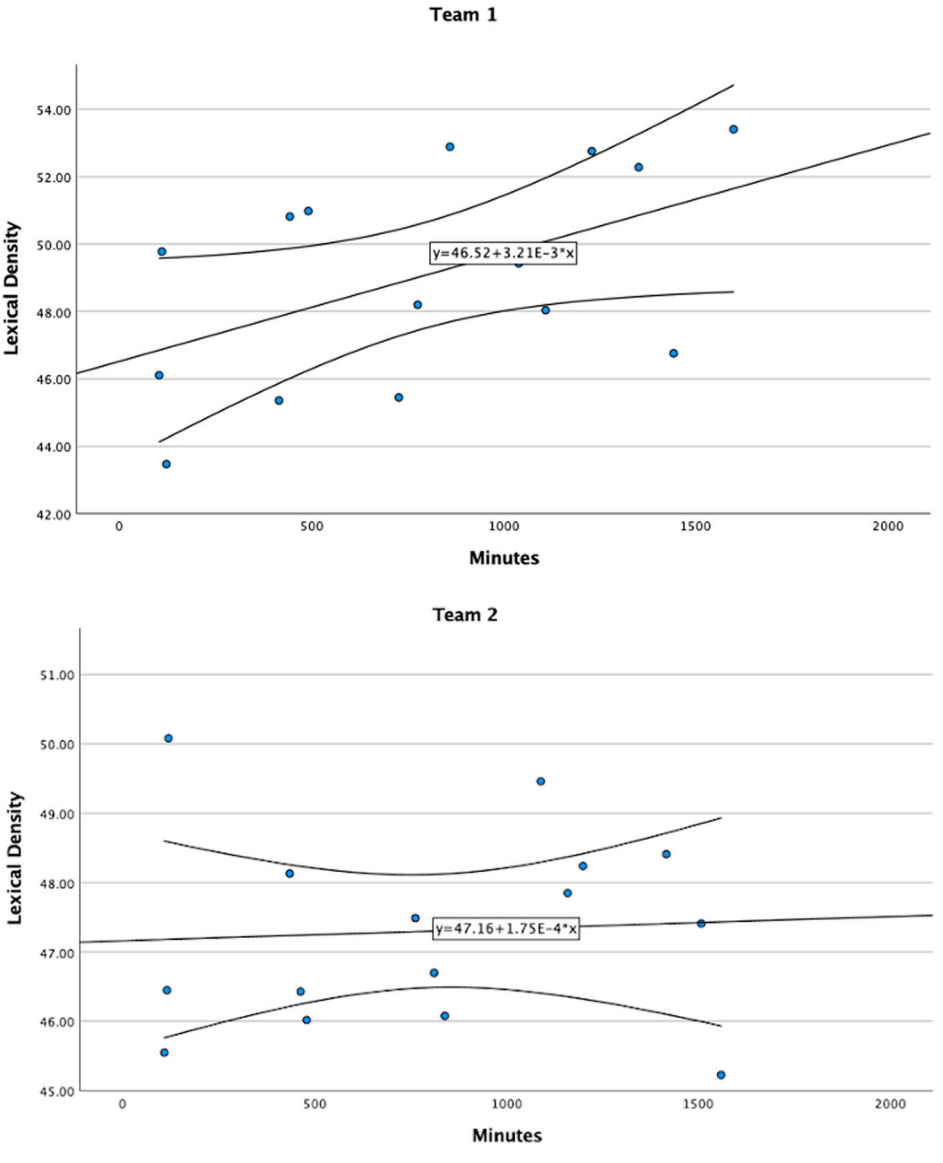

Figure 1. Changes in the average values of lexical density over time.

communicate. On average, the cumulative 95% of the word types used in T1's and T2's spoken output resides in the range of 1k-3k and 1k-4k, respectively. The trendlines for T1 show that word use patterns did not change much in that group. In T2, there was, however, a slight decrease in the percentage of 1k words used over time. This growth in lexical sophistication over time is clear in the data of E from T2, shown in Figure 3. In his spoken production, the proportion of 1k words decreased while that of 3k words increased. E commented explicitly at interview that gameplay allowed him to 'brush up' on his 'rusty' English language skills and regain his English language proficiency (see Excerpt 9 below).

Lexical variation

The results for T1, shown in Figure 4, display an ascending trendline in average D values. This ascending trend is seen notably in MH's data (see Appendix, Figure A.1). Despite this improvement, MH

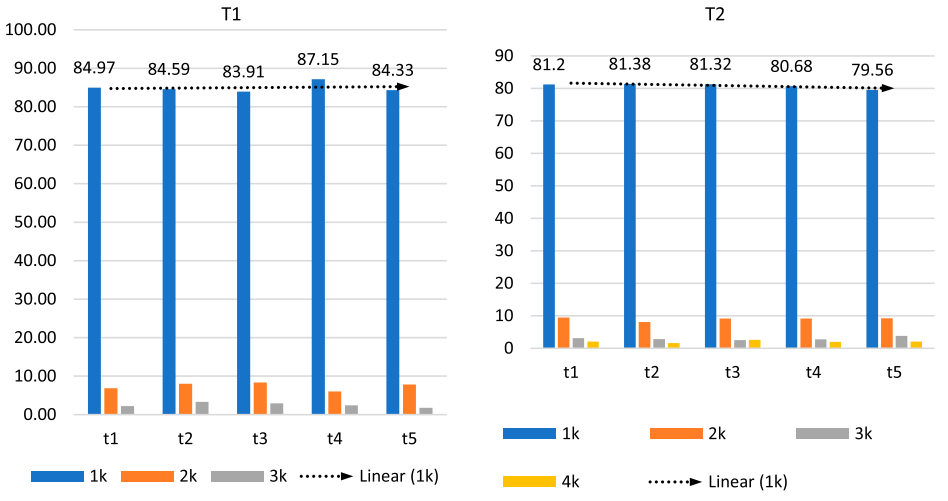

Figure 2. Average percentages of word types in EFL gamers' spoken output.

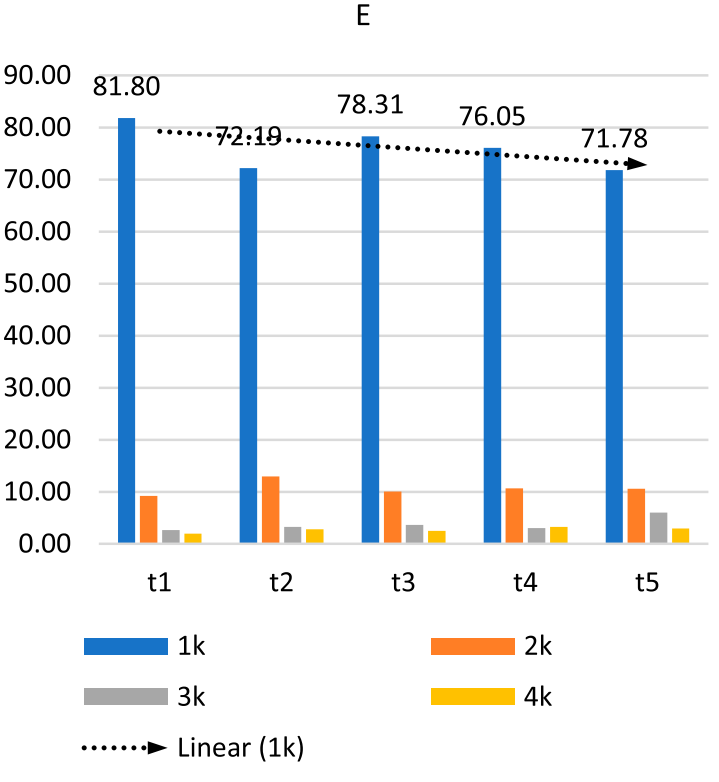

Figure 3. Average percentages of word types in E's spoken output.

believed that 'a limited range of vocabulary' was sufficient for communicating in the game environment (Excerpt 3).

Excerpt 3 (MH's reflection, gaming session 8)

> In this session, we levelled up our characters faster with Nate's assistance as a Healer and M's support as a Tank. This time, I realised that when communicating with a native speaker, it is possible to get your message across with a limited range of vocabulary.

Although MM's lexical variation did not progress as consistently as MH's (see Appendix, Figure A.1), the trendline of average D values in his case shows a slight overall increase, despite a decrease from t2 to t5, indicating that MM expanded the range of vocabulary he used relative to t1. He seems to have developed greater willingness to communicate with Nate. As shown in Excerpt 4, to enjoy this communication, he believed he needed to broaden his range of vocabulary.

Excerpt 4 (MM's reflection, gaming session 9)

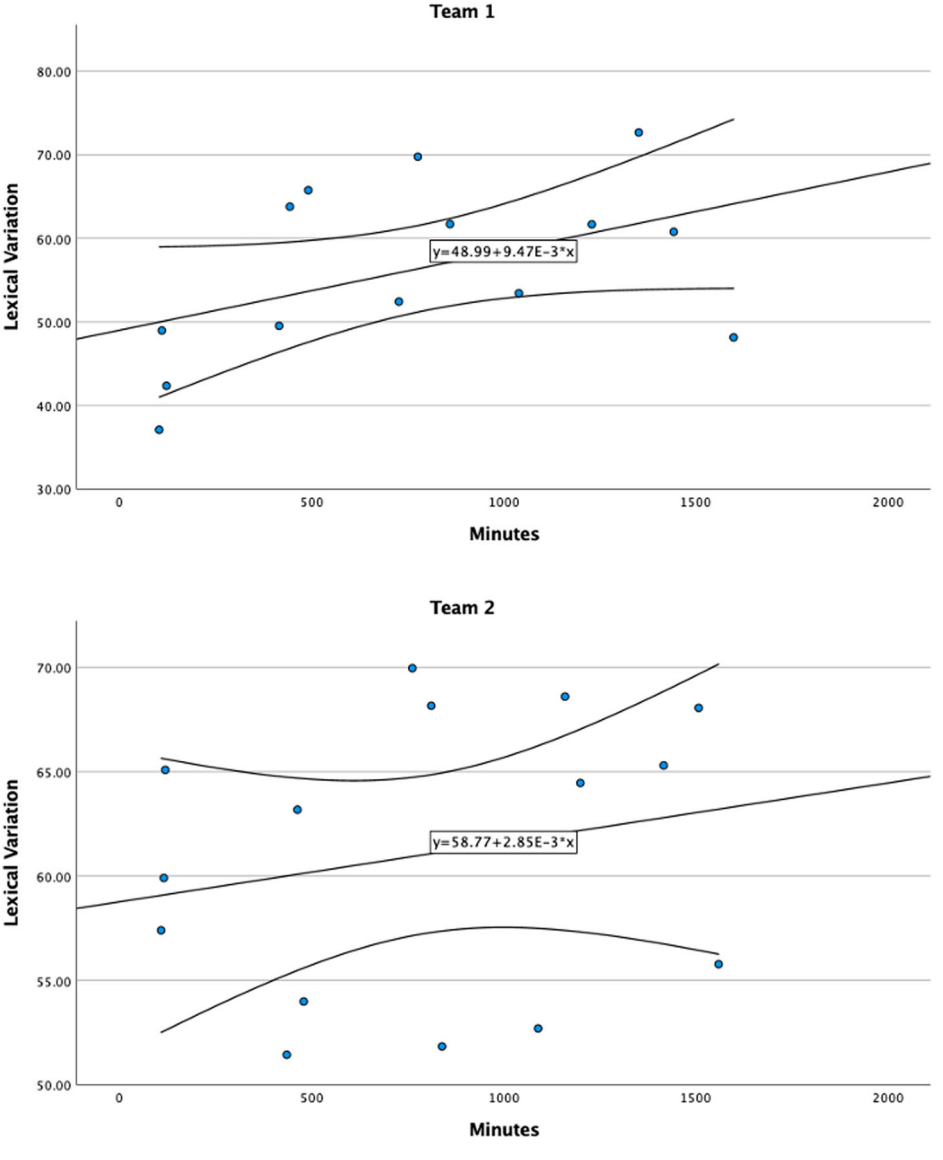

Figure 4. Changes in the average D values over time.

> During the gameplay today, I realised that I need to learn more words because I really enjoy talking with Nate … . Gradually, I feel that my [English] language level is much higher, and I need to learn more words in addition to some simple grammar …

In T2 (Figure 4), D values show some improvements over time but are less prominent than the trend observed in T1. Individual samples (shown in Appendix, Figure A.1) reveal a slight overall increase (participant B), a sharp increase (E), and a slight decrease (M). The slight rise in D values in B's data may reflect his assertion (Excerpt 5) that in-game conversations are more manageable than face-to-face conversations since a gamer requires a more limited range of vocabulary.

Excerpt 5 (B's reflection, gaming session 10)

> In my opinion, conversations in the game environment are more effortless than face-to-face conversations. Because the range of vocabulary you need to take the game forward is more limited … some words and expressions are repeated throughout the gameplay and therefore stick to your mind.

As was the case for lexical sophistication, the clear overall increase in D values in E's utterances may reflect his declared intention to seize the opportunities offered by gameplay to revive his English (see Excerpt 9 below).

Syntactic complexity

We calculated the average number of subordinate clauses per AS-unit for each EFL gamer's data across the five time points. As shown by the trendline in Figure 5, the syntactic complexity of the EFL gamers' speaking performance in T1 did not appear to change much over time. The trendlines for individual data (see Appendix, Figure A.2) show a negligible increase (participant F), an increase (MH), and a decrease (MM). The patterns are a little different for T2, showing small improvements over time. Close examination of individual data reveals ascending trendlines in B's and E's data but a descending trendline in M's data over time (see Appendix, Figure A.2). This may reflect participants' emphasis on the role of vocabulary, not grammar, in in-game communication (see Excerpts 3, 4, and 5 above). Participants noted that simple, succinct, and sometimes telegraphic sentences with a limited range of vocabulary helped them communicate efficiently and effectively during the gameplay. Much less emphasis on grammatical complexity and accuracy may explain why MH acknowledged that his grammar was the least developed after six months of gameplay (see Excerpt 6 below).

Morphosyntactic accuracy

We calculated the percentages of erroneous clauses and AS-units to measure the levels of morphosyntactic accuracy. The change patterns (Figure 6) show a slight decrease in the average percentage of erroneous clauses but a nearly plateaued average trendline for erroneous AS-units in T1. The individual data of erroneous AS-units reveal slight increases in errors in F's and MH's speaking performance but a sharp decrease of errors in MM's (Appendix, Figure A.3). Aligned with these findings are MH's and MM's comments (Excerpts 6, 7, and 8), underlining the significance of meaning over form during in-game communications.

Excerpt 6 (MH's interview 05:04)

> Before all this, I think I have developed more courage to speak in English. My grammar is the least developed, though. I know more words now and have improved my listening and speaking skills.

Excerpt 7 (MM's interview 02:54)

> Unconsciously, a part of your mind—obsessed with some thoughts such as 'I should not speak incorrectly'—gets engaged with the gameplay. Your focus is no longer on your words and sentences [while talking]. Instead, your focus is mainly on the gameplay. As you've noticed, we were more silent on some occasions when we were

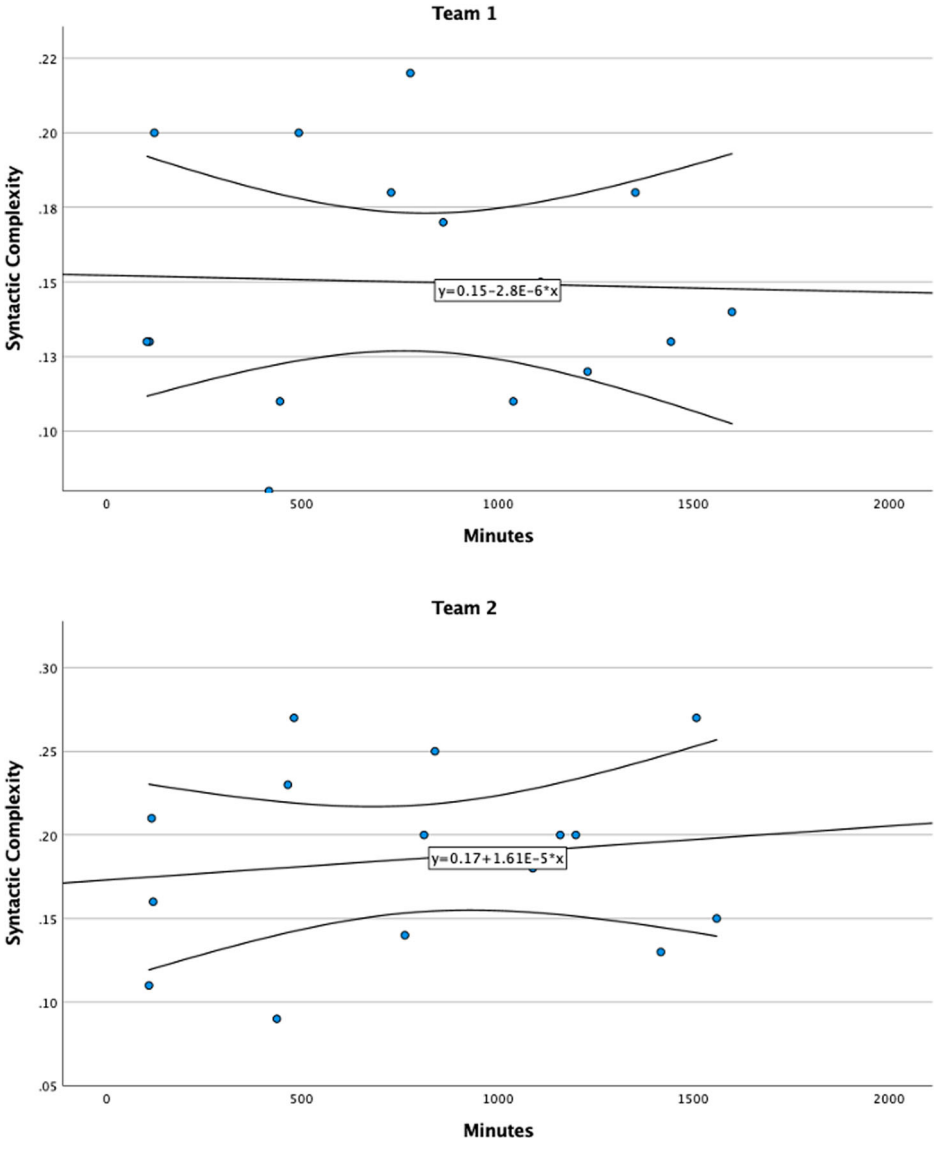

Figure 5. Variations in the average number of subordinate clauses per AS-unit over time.

under pressure and had to make decisions in milliseconds. We were using only the keywords we needed quickly … I didn't care whether I was or wasn't talking correctly. I just wanted to convey my point and ensure that others understood what I meant to say.

Excerpt 8 (MM's interview 10:20)

I also think that I have improved my speaking. I may have some grammatical mistakes, but I can communicate better now. I can speak and get my work done well.

The participants' comments imply that, in their view, morphosyntactic accuracy does not necessarily play a crucial role during in-game communication and that they managed to communicate effectively despite grammatical inaccuracies in their utterances.

The average trendlines for morphosyntactic accuracy in T2 show (Figure 6) similar patterns to T1, i.e. a slight decline in the average number of erroneous clauses and a nearly plateaued

average trendline for erroneous AS-units over time. However, individual data of erroneous AS-units show a possibly more positive picture; that is, descending trendlines in E's and M's data, i.e. fewer errors made, although B's trendline is ascending (See Appendix, Figure A.3). At interview, E commented (Excerpt 9) that in-game conversational exchanges helped him brush up on his 'rusty' English and improve his morphosyntactic accuracy, among other aspects of his English language proficiency.

Excerpt 9 (E's interview 3:10)

In the beginning, my [English] language skills were rusty, as I had not practiced my English for nearly two years. After the first and second sessions [of gameplay], I felt like I was brushing up on my English … If I want to speak English with you, my utterances must be correct grammatically and lexically; that is why I may stutter a little bit.

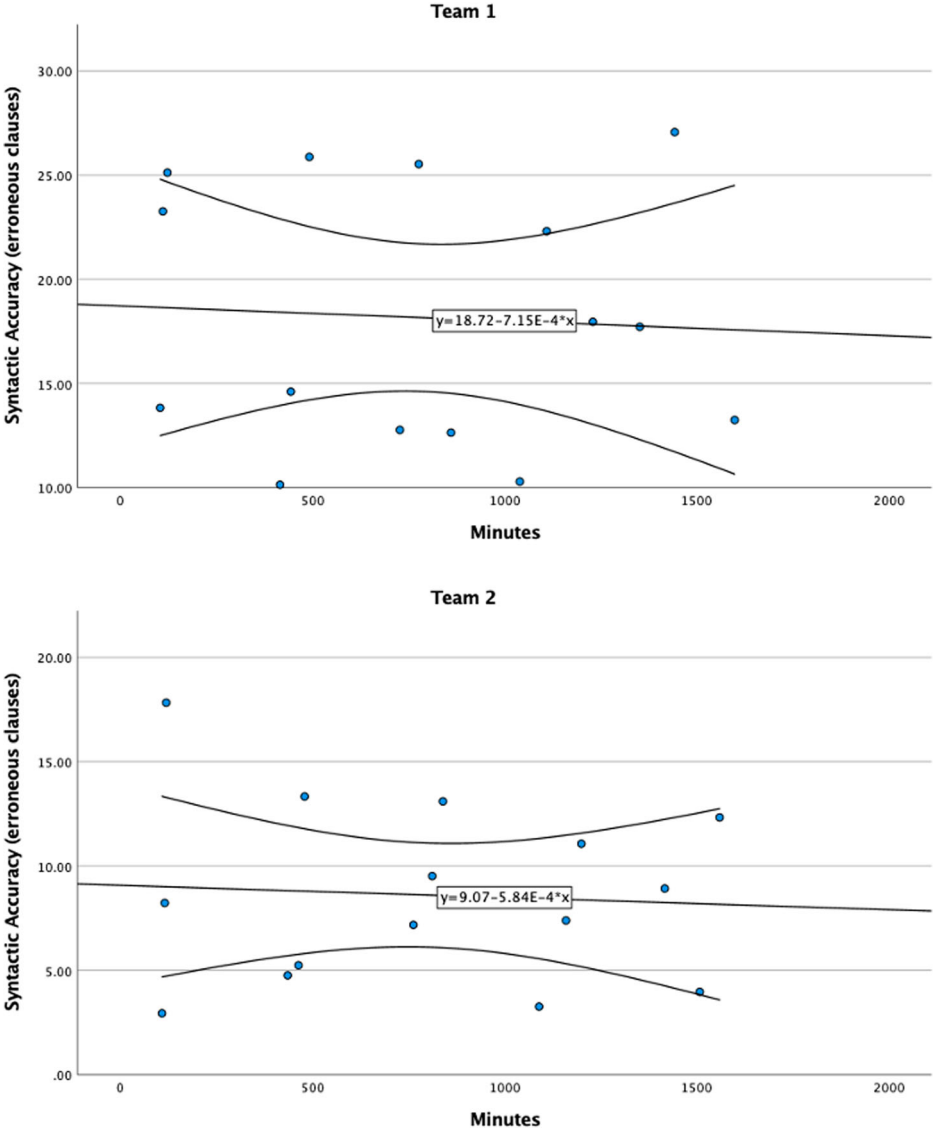

Figure 6. Variations in the morphosyntactic accuracy indices over time.

On the other hand, B's comments at interview (Excerpt 10) may shed light on his pattern of increased morphosyntactic errors, reflecting his gradual shift from accuracy to fluency over time.

Excerpt 10 (B's interview 2:49)

It was difficult at the beginning. I was afraid of speaking [in English] and was super careful not to say something incorrectly. I assumed my grammar had to be perfect and my words had to be flawless. As time passed, I became more relaxed in speaking. As we became friends, I could speak more confidently. I felt Nate understood that I am not a native speaker [of English] and thus may make mistakes and need some time to remember words. Therefore, I experienced less anxiety and developed some self-confidence in speaking.

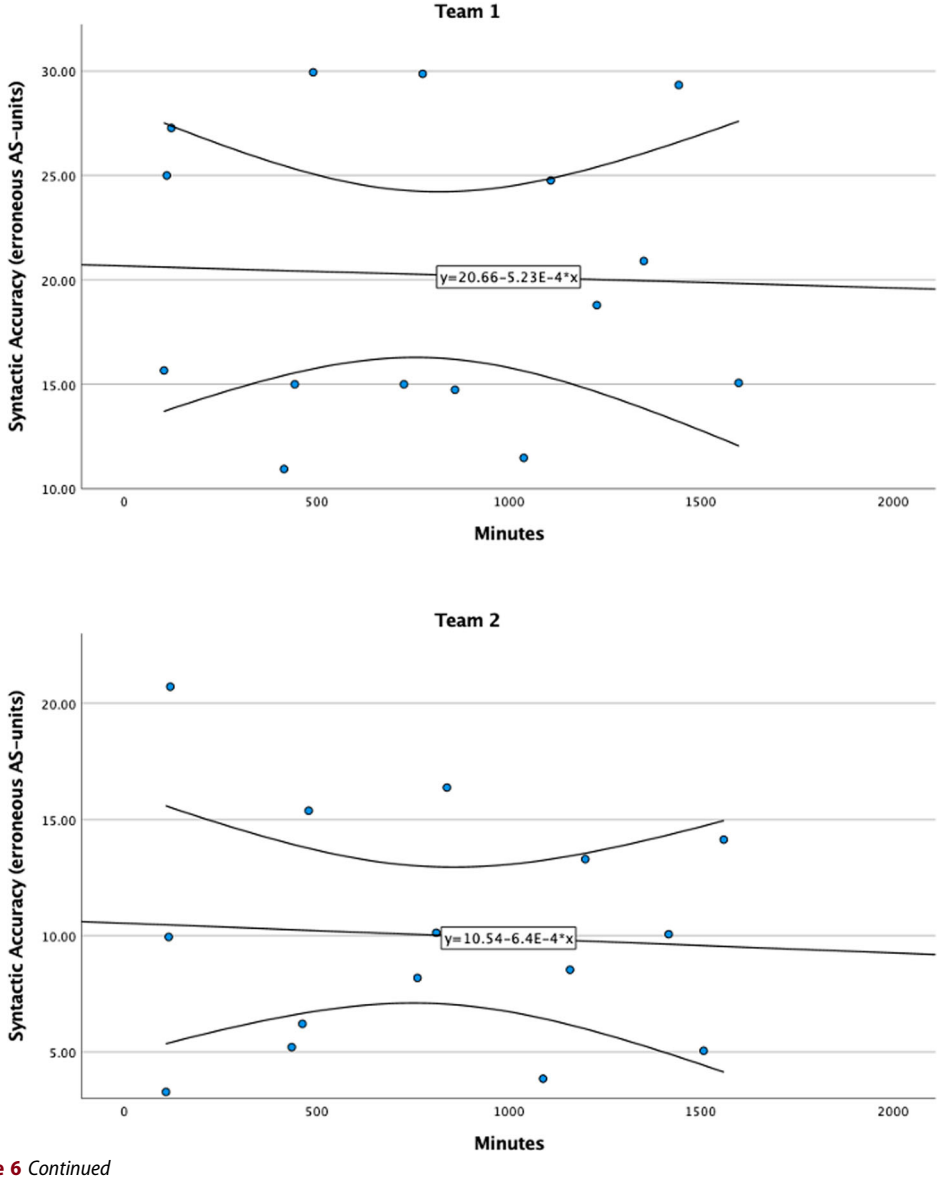

Figure 6 *Continued*

Fluency

Breakdown (dys)fluency

We calculated the average number of silent and filled pauses per AS-unit as well as silent and filled pause ratios to explore the patterns of breakdown (dys)fluency. Overall, the results were broadly positive, reflecting increases in spoken fluency. As shown in Figure 7, T1 had fewer silent pauses per AS-unit, and T2 had fewer filled pauses per AS-unit over time. Figure 8 shows that over time, the gamers in T1 produced shorter silent and filled pauses whileT2 made shorter filled pauses.

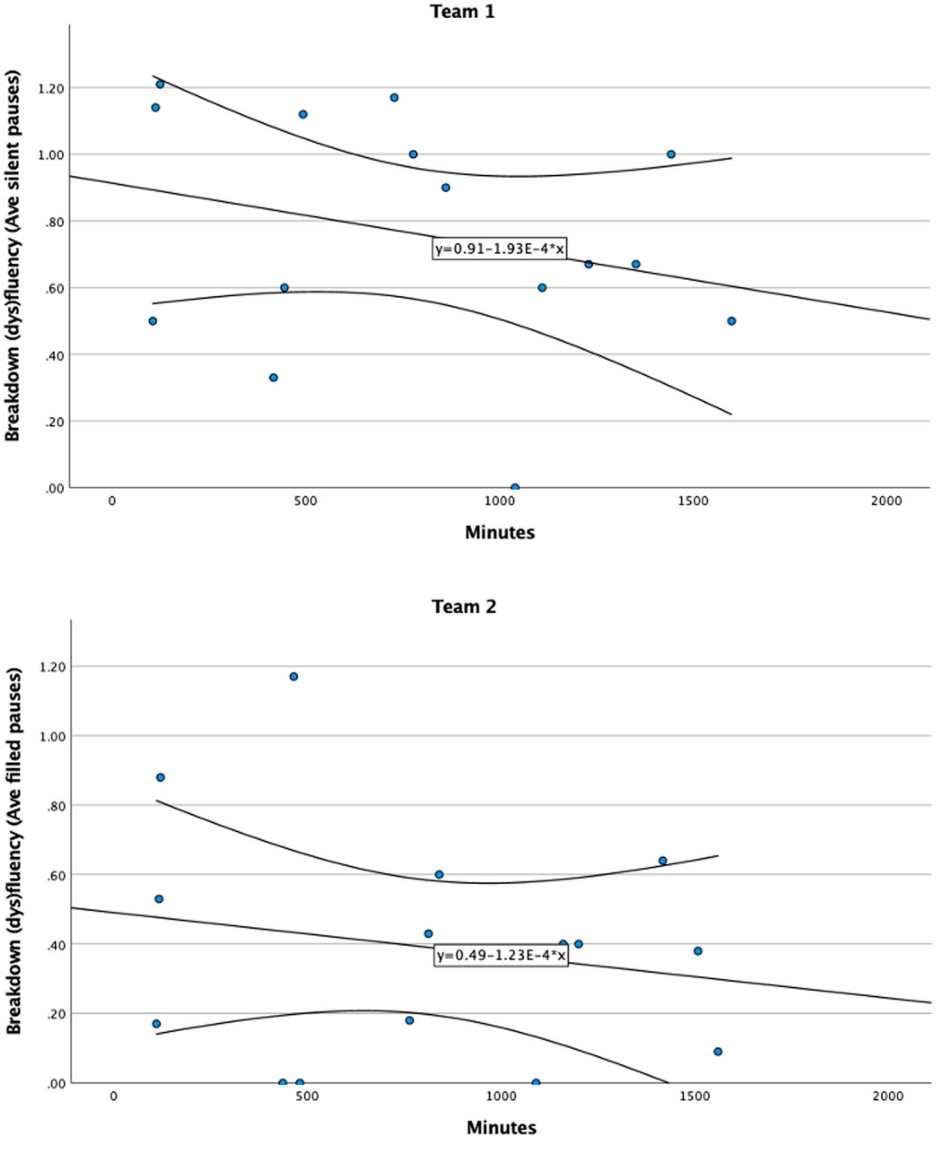

Figure 7. Average number of silent and filled pauses per AS-unit in T1 and T2.

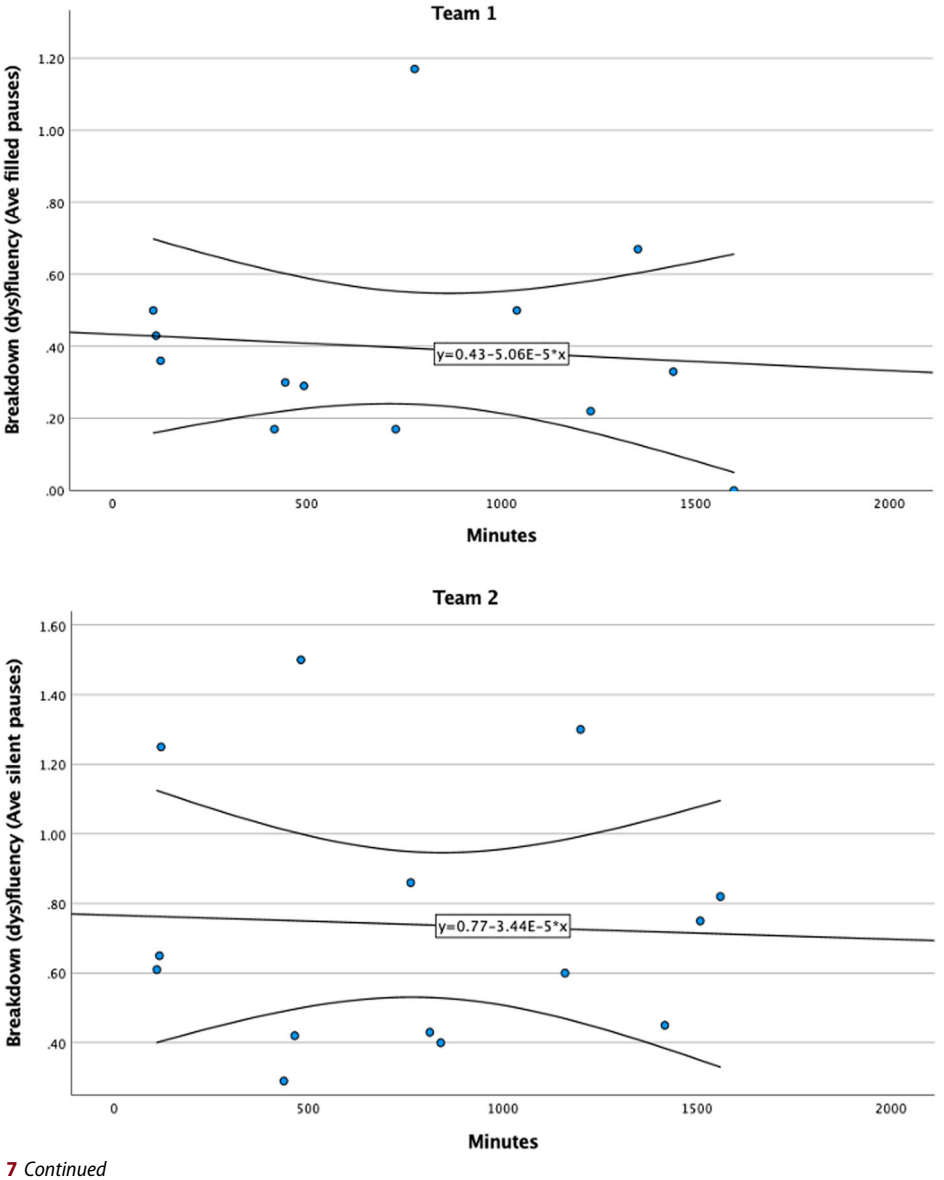

Figure 7 *Continued*

Repair (dys)fluency

As shown in Figure 9, the almost plateaued trendlines in both teams indicate no change over time in the average number of repair (dys)fluency markers per minute. Individual data, however, reveal increasing trends in the number of (dys)fluency markers per minute (e.g. see F in T1 and B in T2, Appendix, Figure A.4).

Speed fluency

Among the three indices of speed fluency (Figure 10), the EFL gamers in T1 improved their phonation time ratio, and those in T2 showed slight improvements in their speech rate per minute. That is, T1 dedicated more of their speaking time to producing language, whereas T2 produced more syllables per minute during their speaking time. Overall, the EFL gamers in both teams improved some aspects of their speaking fluency. These improvements reflect the essence of in-game conversational

exchanges, i.e. effective communication of meaning is required within a restricted time frame (see Excerpts 11, 12, 13 and 14).

Excerpt 11 (M's interview 21:20)

When we were in dungeons, my priority was speaking fast. It was vital for me to think in English. I can say that only 20% to 30% of my mind was engaged with grammar.

Excerpt 12 (B's interview 22:20)

In the beginning, I had to think much more while speaking English. I was thinking about what to say and how to say it. I needed more time to organise my thoughts and structure my sentences. I had to think about every single word. As such, I was talking much less at the beginning. But when we continued, I was talking without thinking

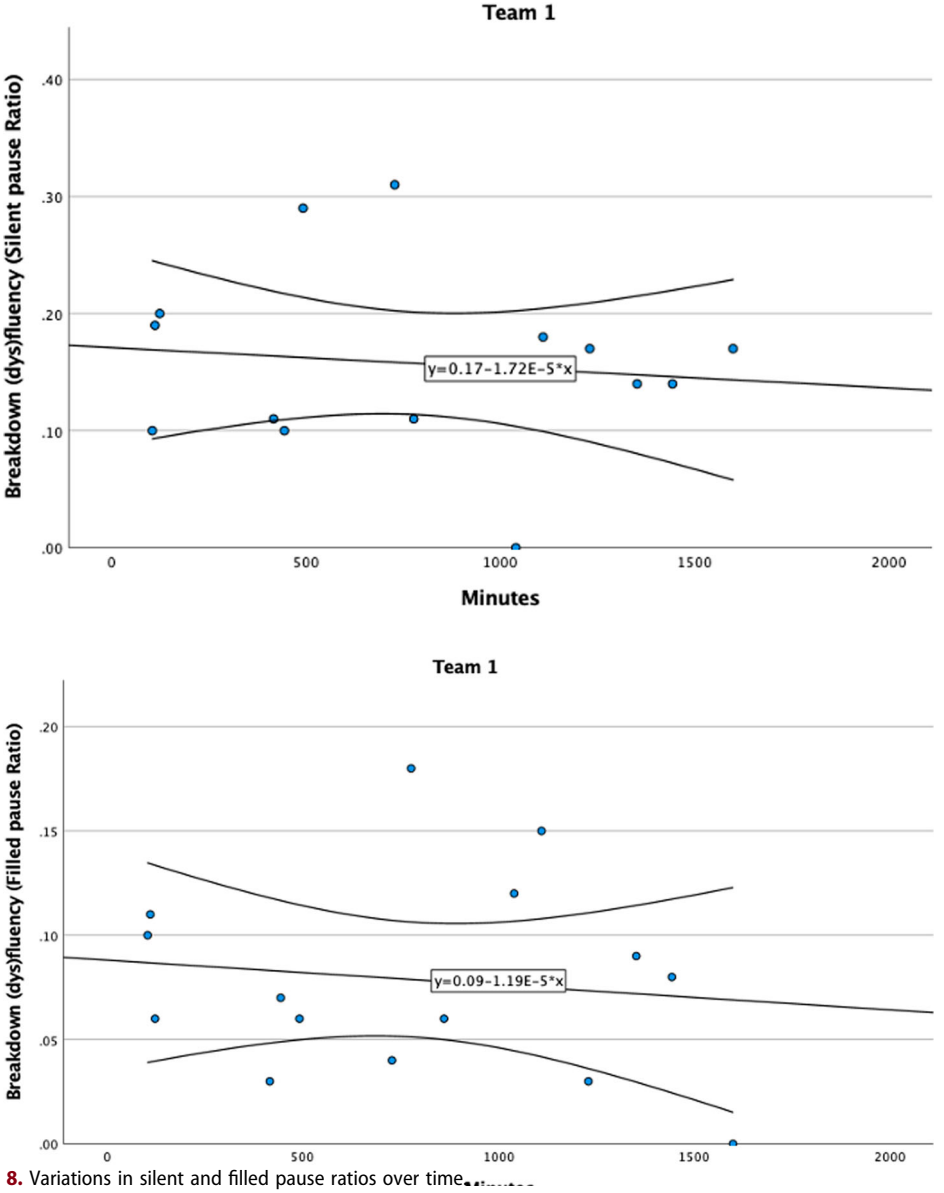

Figure 8. Variations in silent and filled pause ratios over time.

about this stuff, and I was talking much faster. It happened unconsciously. It caused me to speak more confidently and more fluently.

Excerpt 13 (MM's reflection, gaming session 4)

When I was focused on the gameplay, I could speak more comfortably. I had much less focus on pronunciation and grammar, which I think helped me speak more fluently. Of course, I have no idea about the accuracy of my utterances [on such occasions]. I just felt that everyone understood what I was saying.

Excerpt 14. (B's reflection, gaming session 5)

During critical circumstances, I was speaking in English quite subconsciously, and I think more fluently. I was not thinking about or even aware of speaking in English. I could also comprehend [other players' utterances] much more effortlessly. It is surprising.

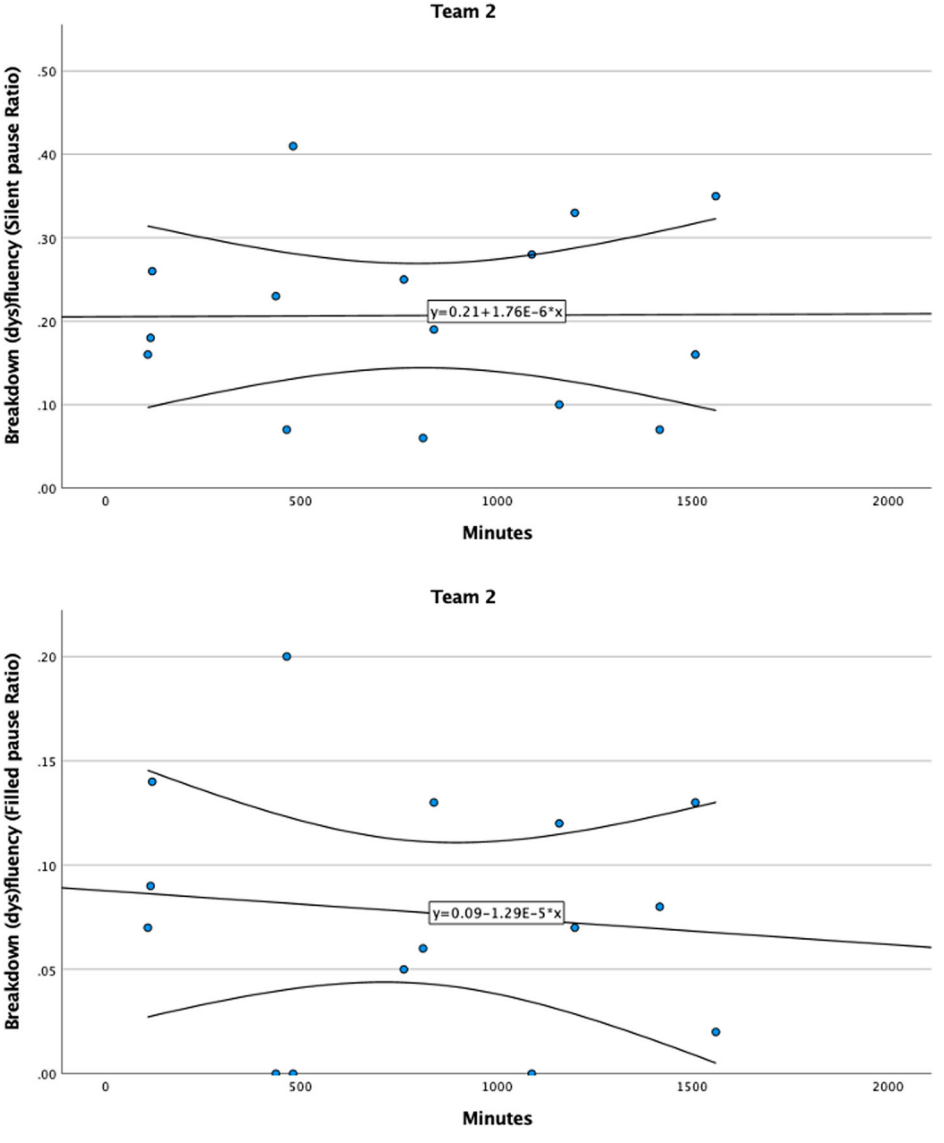

Figure 8 *Continued*

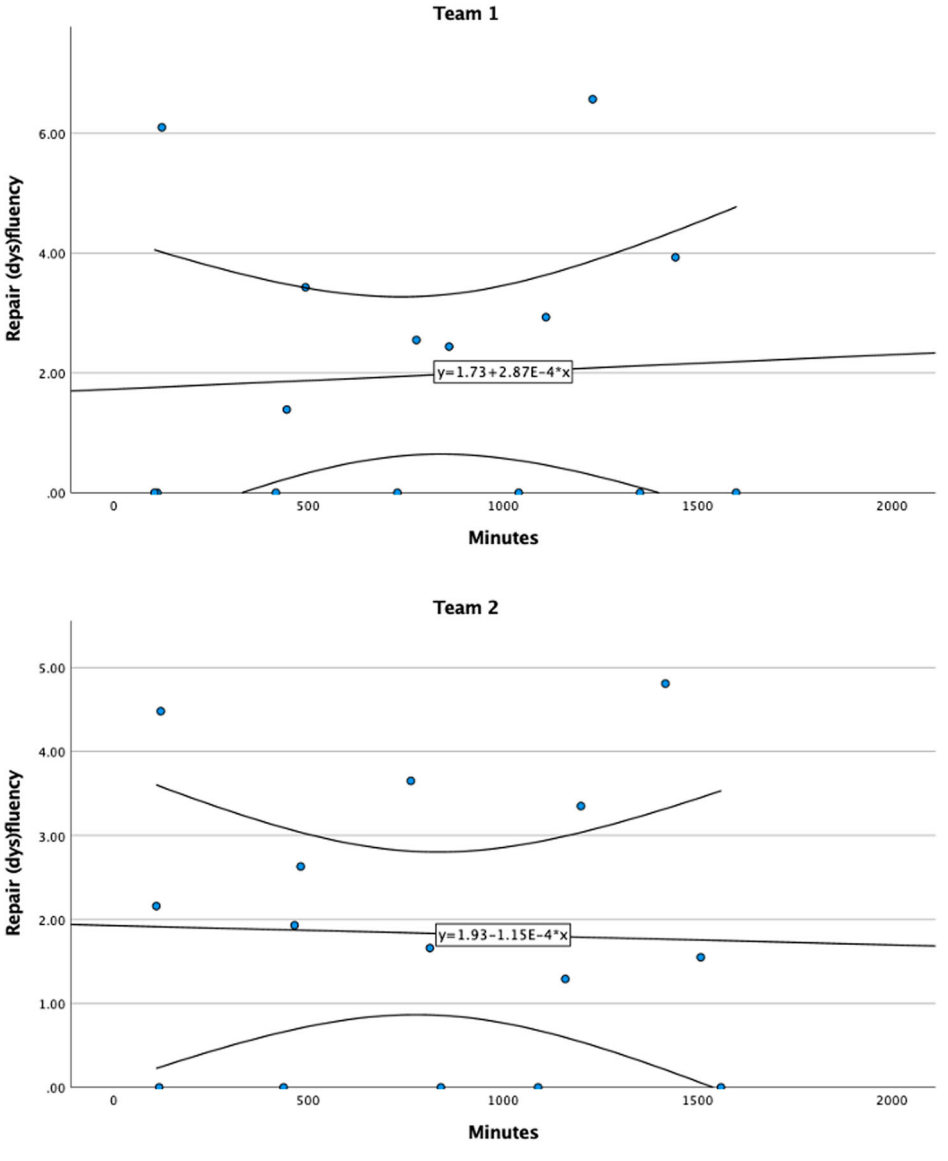

Figure 9. Variations in the average number of (dys)fluency markers per minute.

Discussion

In this research, we sought to explore whether collaborative MMO gaming (with peers and a native English speaker) can improve EFL gamers' speaking performance over time, in terms of complexity, morphosyntactic accuracy, and fluency. To this end, we used 14 measures to gauge CAF levels across five equally distributed data collection points over six months. Overall, the results appeared positive for the EFL gamers on both teams but in different ways (see Table 3).

These indicative findings contradict some scholars' (e.g. Poole and Clarke-Midura 2020; Rankin, Gold, and Gooch 2006) argument that learners should possess at least an intermediate level of L2 proficiency to be able to improve their conversational skills through MMO gaming. For instance, Rankin et al. (2006) found that high-beginner ESL students were cognitively overloaded by the

multiple competencies required to navigate the game, comprehend the information displayed on the screen, and look up unfamiliar vocabulary. In the following sections, we discuss the findings under three broad categories: complexity, morphosyntactic accuracy, and fluency.

Complexity

The EFL gamers' speaking performance in T1 improved in terms of lexical density and lexical variation (Figures 1 and 4). This means they generally produced more informative (or denser) and lexically more varied spoken discourse over time. It can be inferred that T1 grew more efficient at L2 communication

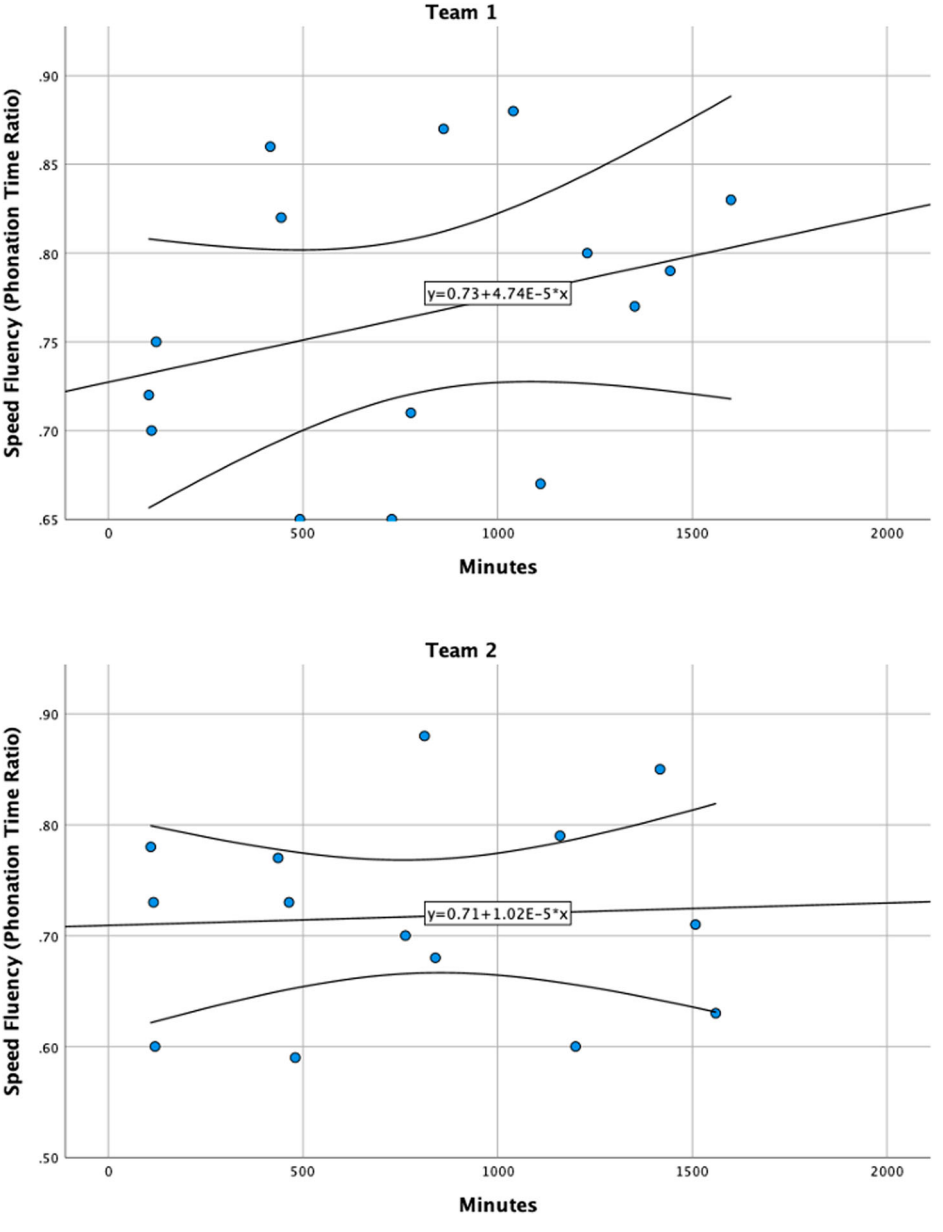

Figure 10. Variations in the indices of speed fluency over time.

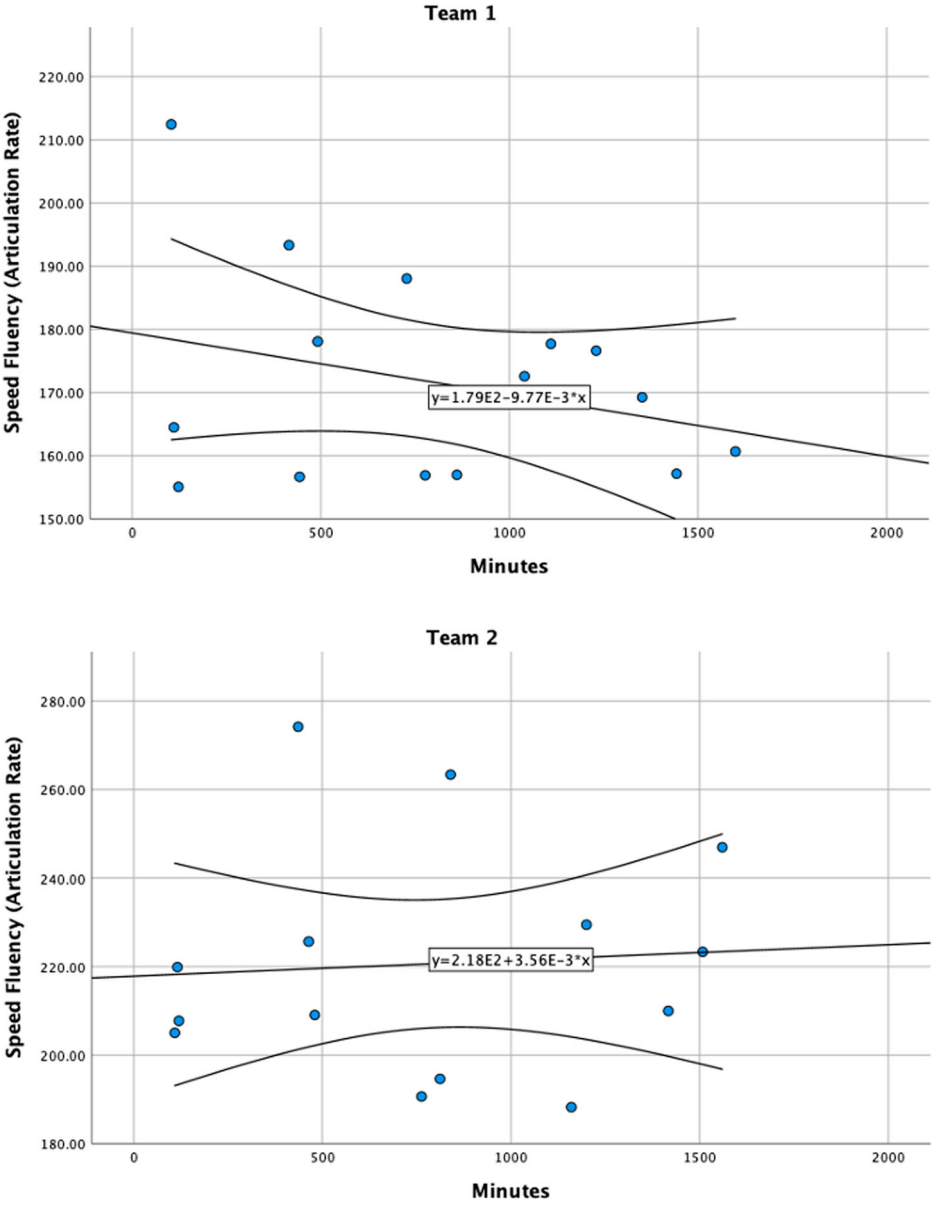

Figure 10 *Continued*

by incorporating more content words into their utterances. However, the lexical sophistication of T1's speaking performance remained the same, indicating that they did not employ relatively more advanced words to communicate over time. T2's speaking performance exhibited no improvements in lexical density (Figure 1) but some advances in lexical sophistication and lexical variation (Figures 2 and 4). As the trendlines in Figure 4 display, T1's spoken output showed greater progress in lexical variation, which could be due to the ascending trend of lexical density in their speaking performance. The correlation between lexical density and lexical diversity (Johansson 2008) suggests that a text with higher lexical density (i.e. higher proportions of content words) can have a higher level of lexical variation (i.e. a wider range of vocabulary). The literature (e.g. Rankin et al. 2006; Sylvén and Sundqvist

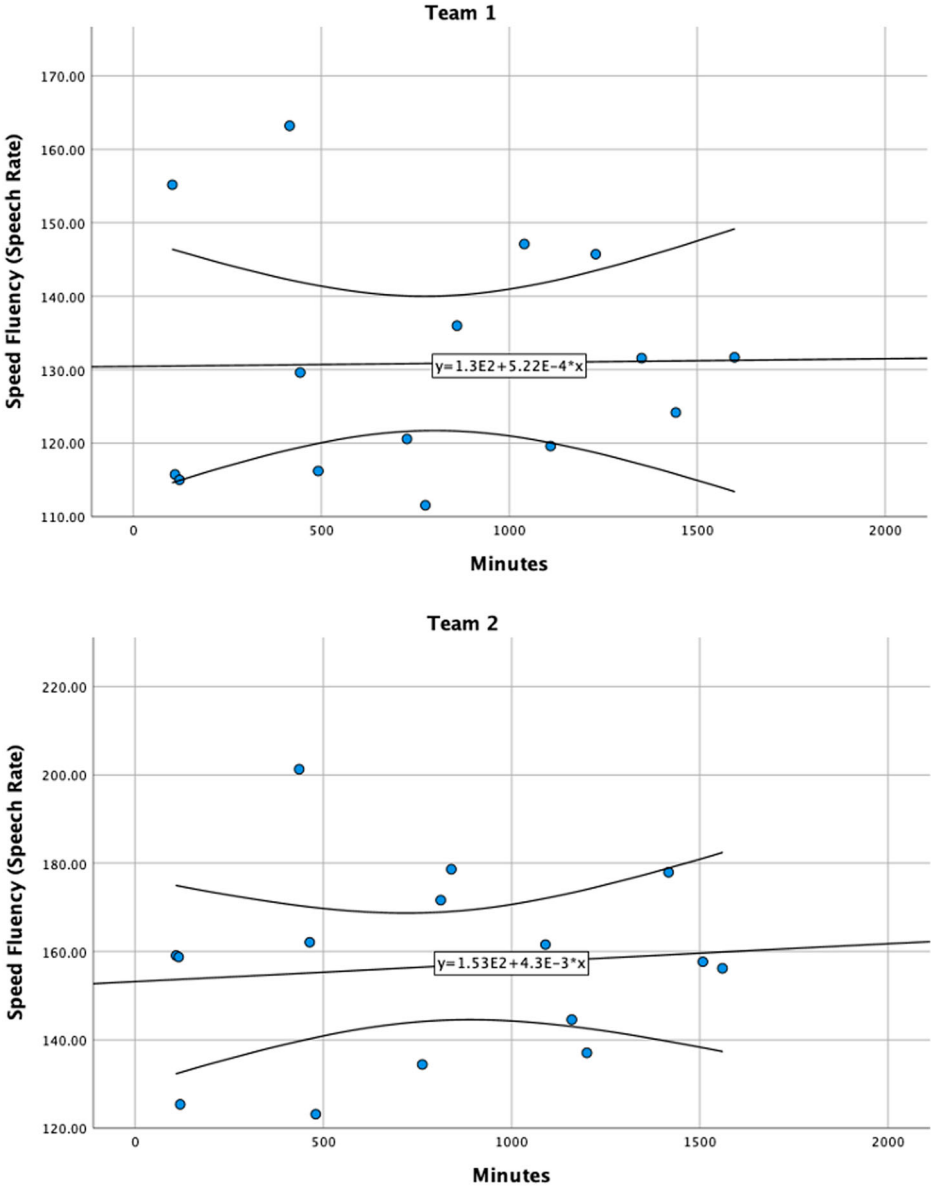

Figure 10 *Continued*

2012) also confirms these results by highlighting that L2 gamers focus primarily on meaning to meet 'the demands for simultaneous communication flow' (Reinders and Wattana 2011: 16) during the gameplay. These findings are also aligned with the EFL gamers' knowledge of the dynamics of in-game communication, which prioritise focus on meaning, necessitate speed in interactions, and promote efficiency in language use during the gameplay. The improvements of T1's speaking performance (in two indices of lexical complexity) contradict Rankin, Gold, and Gooch (2006) findings that intermediate and advanced ESL students could improve their vocabulary more than high-level beginners. Rankin, Gold, and Gooch (2006) concluded that playing *Ever Quest II* failed to provide adequate language learning support for lower-level ESL students.

Table 3. An overview of improvements in CAF indices over time.

		T1	T2
Lexical complexity	Lexical density	✓	
	Lexical sophistication		✓
	Lexical variation	✓	✓
Syntactic complexity	Ave number of subordinate clauses per AS-unit		✓
Morphosyntactic accuracy	% of erroneous clauses	✓	✓
	% of erroneous AS-units		
Breakdown (dys)fluency	Ave number of silent pauses per AS-unit	✓	
	Ave number of filled pauses per AS-unit		✓
	Silent pause ratio	✓	
	Filled pause ratio	✓	✓
Repair (dys)fluency	Repair (dys)fluency		
Speed fluency	Phonation time ratio	✓	
	Articulation rate per minute		
	Speech rate per minute		✓

The syntactic complexity measures revealed no change for T1 but minor improvements for T2 (Figure 5). These results should be interpreted cautiously as not all communication contexts necessarily demand syntactically complex discourse. That is, on some communication occasions, simpler AS-units are more efficient and signify higher levels of communicative competence (Ferrari 2012). As Jabbari and Eslami (2019: 104) have noted, the characteristics of in-game interactions (e.g. task-based, goal-driven, meaning-oriented, multimodal, multilateral, and time-sensitive) promote a style of verbal communication that is mostly 'unorthodox in language form, succinct in nature, and innovative in style'. Therefore, the results should be interpreted in light of the fact that 'not only is the syntactic complexity construct multi-faceted, but it is also problematic to assume that it grows in a steadily linear way, or even that it grows over time altogether' (Ferrari 2012: 283). Ortega (2003: 494) has warned that 'syntactic complexity metrics would be misapplied if they were to be used as absolute developmental indices or as direct indices of language ability'. L2 development is a multi-faceted, non-linear process that is too complex to be measured precisely and explored sufficiently by implementing global measures alone. Having underlined the socio-pragmatic and sociolinguistic aspects of L2 proficiency, Ortega (2003: 494) argued that '[p]rogress in a learner's language ability for use *may* include syntactic complexification, but it also entails the development of discourse and sociolinguistic repertoires that the language user can adapt appropriately to particular communication demands' (*emphasis added*).

Morphosyntactic accuracy

The results appeared similar for the EFL gamers in T1 and T2. Their speaking performance generally grew more accurate, containing fewer erroneous clauses over time. However, the percentages of erroneous AS-units in their utterances remained nearly the same over time (Figure 6). These results differ slightly from what we know from the literature. For example, Reinders and Wattana (2011), who conducted their study with very similar participants to those in T1, reported different results. Unlike the current research design, where participants engaged in a totally authentic *WOW* experience, Reinders and Wattana (2011) extended *Ragnarok Online*[6] by adding three new quests, instructed the participants to collaborate in the gameplay, and encouraged them to improve their English by using it during the gameplay. They found that in-game L2 interactions increased the quantity of the students' discourse but did not enhance the accuracy and complexity of their utterances in either text-based or voice-based chat groups. Meanwhile, Rankin et al. (2009) found that the ESL students who attended class instruction gained higher post-test scores for appropriate vocabulary use in sentences than those who played *EverQuest II* on their own or with NESs.

It is noteworthy that the communication norms and priorities governing the gameplay should be considered when interpreting these findings. Producing an inaccurate 'linguistic form (i.e. grammar,

vocabulary, spelling, discourse, or pronunciation)' (Ellis, Basturkmen, and Loewen 2001: 294) does not always denote an inadequate or lack of L2 knowledge. The social and cultural dynamics of a communication setting and the digital medium of communication that dominates are among multiple factors that can promote an idiosyncratic vernacular that sounds more appropriate than a standard and orthodox form of language. As such, it is not always straightforward to determine whether an incorrect L2 usage signifies an 'error,' a 'mistake' (Brown 2000), or a conversation style that legitimises conscious deviations from the structural rules of the target language. Rama et al. (2012: 331–2), for instance, observed that '[…] even native speaker gamers use abbreviated and orthographically and stylistically *non-standard language*, in addition to simple typo-graphical errors, often because they are attending to multiple and simultaneous activities within the game' (*emphasis added*). They discovered that cultural norms ruling in-game communication prioritised function over form. They explained that the gaming environment favoured sociolinguistic and strategic competencies (Canale and Swain 1980), which emphasise 'contextualised meaning rather than grammatical and lexical correctness of standard language forms' (Rama et al. 2012: 330). As such, every successful move in the game warrants purposeful L2 use for (re)(co)constructing and communicating meaning (Steinkuehler 2007; Jabbari and Eslami 2023). In such communication contexts, pragmatic appropriateness becomes more critical than grammatical accuracy (Godwin-Jones 2014). Accordingly, the accuracy of linguistic form matters only if it contributes to the underlying processes of meaning-making endeavours. Aligned with this argument are some scholars' (e.g. Reinders and Wattana 2011; Reinders and Wattan 2015a) findings that L2 interactions during gameplay helped ESL learners develop a set of communicative skills/strategies and encouraged them to employ various discourse functions.

Similarly, we observed that the EFL gamers were actively involved in the co-construction and communication of meaning by investing more in content words (see Excerpts 2, 7, 8, 11, and 13). As reported in a complementary study (Jabbari and Eslami 2023), the EFL gamers also implemented other communicative strategies in their meaning-making efforts. One of these strategies was drawing on the game's multimodal contextual features (e.g. icons and symbols, sounds and music, shapes and colours, as well as playing and non-playing characters' gestures, movements and actions). It is thus plausible to argue that L2 gamers attend to the morphosyntactic accuracy of their utterances as long as it contributes to their meaning-making efforts. Otherwise, gamers combine innovatively various components of language with the game's semantic and semiotic affordances (Jabbari and Eslami 2023) to develop an effective communication medium during the gameplay. In the same vein is Ibrahim's (2022: 1181) argument that MMOG players 'are more likely to observe and learn linguistic constructions that are *instrumental* to guiding in-game action' (*emphasis added*).

Fluency

Both teams made some progress in breakdown (dys)fluency and speed fluency but not in repair (dys)fluency (see Table 3). Concerning breakdown (dys)fluency, the EFL gamers in T1 grew more fluent, having fewer silent pauses per AS-unit and producing shorter silent and filled pauses over time. The EFL gamers in T2 also showed increases in speaking fluency with fewer filled pauses per AS-unit and shorter filled pauses over time. Among speed fluency measures, T1 showed some improvements in phonation time ratio and T2 in speech rate per minute. In other words, T1 managed to produce higher proportions of language during their speaking time, and T2 managed to produce more syllables in their speaking time per minute. We can argue that collaboration under time pressure to complete in-game tasks pushed the EFL gamers to produce spontaneous language (Swain 2000) and develop the automaticity of their speaking performance (Reinhardt 2019). These results confirm EFL gamers' claims in the current research and other studies (e.g. Li, Peterson, and Wang 2022; Peterson 2012a, 2012b; Reinders and Wattana 2015b)

that MMO gaming helped them improve their fluency and discourse management strategies in the target language.

The EFL gamers' improvements in fluency can also be explained by drawing on the dynamics of game-oriented communication and the socio-affective ambiance of the game environment (Horowitz 2019; Reinders and Wattana 2015b). As the EFL gamers themselves explained, they prioritised meaning-making and strived for speed and efficiency in their L2 use during game-oriented communications (see Excerpts 2, 7, 8, 11 and 13). They had to manage a trade-off mainly between meaning and form rather than among the threefold dimensions of speaking performance (i.e. CAF). The EFL gamers also highlighted the development of an in-game 'community of practice' (Wenger 1998), which helped with the establishment of a positive socio-affective environment in the game (see Excerpts 3, 6 and 10). Having worked in tandem, these features seem to have liberated the EFL gamers from hindering L2-related affective states (e.g. fear of speaking due to lack of self-confidence and L2 anxiety) and instilled in them the mindset that attending to L2 rules is important only when it contributes to the communication of meaning. As evidenced in the EFL gamers' comments, their positive socio-affective experiences while developing a more pragmatic mindset focusing on L2 use rather than accuracy helped them reach an immersed state of 'flow' (Csikszentmihalyi 1990), which enhanced automaticity in their spoken interactions. They became so absorbed and engrossed in highly interactive, goal-oriented, and fun activities in the game that they no longer felt self-conscious about using the target language (see Excerpts 7, 13 and 14). It is analogous to entering a trance-like state of complete absorption and effortless concentration where they used the L2 unconsciously, i.e. without being metacognitively overloaded by or even conscious of the formal constructing components of the target language.

It should be noted, though, that 'not all MMOGs are designed and played equally' (Reinhardt 2021: 71). Therefore, not all MMOGs will yield similar results. MMOGs have different *game design mechanics*, i.e. 'feature[s] designed into a game operationalised by a game rule that affords and constrains a given type of gameplay behaviour' (Reinhardt 2021: 74–75). They also have different *gameplay dynamics*, which 'emerge out of the interaction of player behaviour with mechanics, and can be thought of actions that are possible' (Reinhardt 2021: 75). In other words, in a gameplay session, an MMOG's designed features (e.g. storylines, rules, narratives, challenges, and goals) constantly interact with player behaviour. These interactions create unique gameplay dynamics, e.g. the way L2 gamers interact (with the game environment and other gamers) and (co)(re)construct and communicate meaning (using multiple modes of communication), which determine how L2 gamers develop their L2 skills with impacts on accuracy, fluency and complexity. As Reinhardt (2021: 70) elucidated, even the gameplay dynamics vary each time a single game is played:

> While a title usually shares some common mechanics with other titles in its genre, a single game title is designed as a unique combination of mechanics, and each time it is played, different mechanics interact with player action to result in different dynamics, which may include engagement, social interaction, language use and language learning.

Therefore, we need to be cautious in generalising the implications of the current research, which focused on EFL learning in the context of *WoW*, to other similar EFL learning contexts, i.e. with other MMOGs in the same genre and with similar configurations of game players.

It is also worth noting that game design mechanics and gameplay dynamics (Reinhardt 2021) determine in-game task conditions and characteristics, which can, in turn, impact L2 gamers' speaking performance in terms of complexity, accuracy, and fluency. Considering the changing dynamics of in-game task conditions and characteristics, it could also be argued that not all underlying principles of existing CAF theories, e.g. Skehan's (2014) Limited Attention Capacity (LAC) hypothesis, can account for L2 gamers' linguistic behaviour in MMOGs. For example, Bui and Skehan (2018: 3–4) explained that '[f]rom the LAC point of view, additional support during the task (e.g. sufficient online planning time or visual aids) would usually lead to improvement in terms of complexity, accuracy, or both at the cost of fluency'. However, the CAF patterns that emerged in this research do not

align fully with this principle. All the game-mediated tasks were embedded in a 3-D setting with numerous multimodal features (e.g. texts, sounds and music, shapes and colours, icons and symbols, as well as playing and non-playing characters' utterances, gestures, movements, and actions). Despite all these verbal and non-verbal (i.e. audio-visual) supports, we witnessed EFL gamers in both teams improving various aspects of their speaking complexity and accuracy but not at the cost of fluency (see Table 3). New CAF theories need to account for the complex dynamics of L2 gamers' linguistic behaviour during MMO gameplay.

Conclusion

As noted earlier, the existing literature on the affordances of commercial-off-the-shelf MMOGs for L2 learning is focused predominantly on text-based interactions during gameplay. Very few studies have investigated L2 gamers' speaking performance and only general accounts of L2 gamers' improvement have been provided. These accounts are mostly based on participants' perceptions rather than detailed measurements of their performance during gameplay. Furthermore, the studies that have measured L2 gamers' speaking performance have assessed speaking without considering its distinct but interrelated subcomponents (i.e. complexity, accuracy, and fluency).

To address this gap, we investigated whether playing an MMORPG in L2 English led to changes in EFL gamers' speaking performance over time. We adopted a usage-based approach to language acquisition to scrutinise any variations in fourteen CAF indices across five equally distributed data collection points over six months. Overall, the results revealed that collaborative playing of WoW in L2 English with peers and an NES led to improvements in some linguistic aspects of the EFL gamers' speaking performance. These improvements, however, varied in type and magnitude across the lower- and higher-proficiency EFL gamers. These results support the notion of 'Naturalistic CALL' (Chik 2013), which emphasises the value of informal language learning in digital communication settings.

The affordances of MMOGs for language learning should be evaluated in light of the primary purpose for which these games are designed; that is, to provide gamers with an ultimate level of enjoyment and fulfilment, namely, 'game flow' (Godwin-Jones 2014: 11). Their primary purpose informs their design features (e.g. storyline, rules, narratives, challenges, and goals) with unique configurations. Dynamic interactions between game design features and gamer-associated variables (e.g. gaming style, motivation, and experience) establish MMOGs' emotional, social, cultural, and semiotic ecologies, which afford different opportunities for L2 socialisation and learning. As this research has shown, using L2 as a medium of communication in the game environment helped EFL gamers improve some but not all linguistic aspects of their speaking performance. This is arguably because COTS MMOGs, in general, and WoW, in particular, are not designed to demand or even encourage a syntactically complex and flawless discourse for in-game communication. Instead, they afford ample opportunities for authentically contextualised L2 use for (co)(re)constructing meaning in a semiotically rich, socially supportive, and emotionally positive environment. By seizing these opportunities, L2 gamers can transform into self-confident and communicatively competent L2 users.

As with many research endeavours, this case study has limitations. As noted, due to the variability in game design mechanics and gameplay dynamics, the findings cannot be generalised to similar EFL learning contexts. Moreover, the number of participants was limited due to circumstances outside the researchers' control. In addition, although the researchers made considerable efforts to implement triangulation by collecting and interpreting data from multiple sources (including participants' game journals and post-study semi-structured interviews), these data sources are subject to well-known limitations on learner self-reporting.

As Godwin-Jones (2014: 11) stated, 'Particularly helpful would be studies that seek to identify what particular user behaviors, game elements, and game resources seem to be the most promising for language learning'. This statement implies that L2 gaming does not guarantee L2 development.

Like in any other learning context, an optimal condition should exist for L2 learning. The main question is which essential variables must interact to generate this optimal condition in MMOGs. What gameplay dynamics should be in place to yield a successful L2 learning experience through playing commercially developed MMOGs? Future research should be inspired by Complexity and Dynamic Systems Theory (Larsen-Freeman and Cameron 2008) and invest in developing a comprehensive model that can explain – based on empirically determined significant paths among a multitude of different variables – how and to what extent the interactions among learner, game, and context-associated variables should work in tandem to facilitate L2 learning through L2 gaming.

Notes

1. MMOGs are not monolithic. They have a variety of different themes, features, in-game cultures, and a unique combination of mechanics.
2. https://www.blizzard.com
3. BNC is a 100-million-word collection of samples (of written and spoken language) from a wide range of sources. COCA is composed of more than 450 million words from more than 160,000 texts. Visit http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf for more information
4. Other measures of lexical variation include TTR, mean segmental TTR, corrected TTR, root TTR, bilogarithmic TTR and the Uber Index
5. Designed at Carnegie Mellon University, CLAN is accessible through the *TalkBank* community, the largest open repository of spoken language data (http://talkbank.org/).
6. Ragnarok Online is one of the most popular MMORPG in Thailand.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Nasser Jabbari http://orcid.org/0000-0001-5715-2265
Mark Peterson http://orcid.org/0000-0002-3497-0203

References

Analyze My Writing. n.d. Anayze my Writing. Accessed June 1, 2023. http://www.analyzemywriting.com/index.html.
Behrens, H. 2009. Usage-based and emergentist approaches to language acquisition. *Linguistics* 47, no. 2: 383–411.
Boersma, P. and D. Weenink. 2013. PRAAT: Doing phonetics by computer (Version 6.1.16). http://www.praat.org/.
Brown, H.D. 2000. *Principles of Language Learning and Teaching*. White Plains, NY: Pearson Longman.
Bui, G. and P. Skehan. 2018. Complexity, accuracy, and fluency. In *The TESOL Encyclopedia of English Language Teaching*, ed. J. I. Liontas, 1–7. Hoboken, New Jersey: John Wiley and Sons.
Bulté, B., and A. Housen. 2012. Defining and operationalising L2 complexity. In Dimensions of L2 Performance and Proficiency: Investigating Complexity, Accuracy and Fluency in SLA, eds. B. Bulté and A. Housen, 21–46. Amsterdam/Philadelphia: Benjamins.
Canale, M. and M. Swain. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1, no. 1: 1–47.
Chapelle, C.A. 2008. Technology and second language acquisition. *Annual Review of Applied Linguistics* 27: 98–114.
Chik, A. 2013. Naturalistic CALL and digital gaming. *TESOL Quarterly* 47, no. 4: 834–39.
Chik, A. 2014. Digital gaming and language learning: autonomy and community. *Language Learning and Technology* 18, no. 2: 85–100.
Chotipaktanasook, N. and H. Reinders. 2018. A massively multiplayer online role-playing game and its effects on interaction in the second language: play, interact, and learn. In *Handbook of Research on Integrating Technology Into Contemporary Language Learning and Teaching*, eds. B. Zou and M. Thomas, 424–46. Hershey, PA: IGI Global.
Cobb, T. 2020. *Compleat Lexical Tutor v.8.3*. Accessed June 30, 2020. https://www.lextutor.ca/.
Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, no. 1: 37–46.
Csikszentmihalyi, M. 1990. *Flow: The Psychology of Optical Experience*. New York: Harper Perrennial.

Dixon, D.H. and M.A. Christison. 2021. L2 gamers' use of learning and communication strategies in massively multiplayer online games (MMOs): an analysis of L2 interaction in virtual online environments. In *CALL Theory Applications for Online TESOL Education*, eds. K. Kelch, P. Byun, S. Safavi and S. Cervantes, 296–321. Hershey, PA: IGI Global.
Dixon, D.H., T. Dixon and E. Jordan. 2022. Second language (L2) gains through digital game-based language learning (DGBLL): a meta-analysis. *Language Learning and Technology* 26, no. 1: 1–25.
Ellis, R. 1994. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
Ellis, N. 2002. Reflections on frequency effects in language processing. *Studies in Second Language Acquisition* 24: 297–339.
Ellis, R., H. Basturkmen and S. Loewen. 2001. Learner uptake in communicative ESL lessons. *Language Learning* 51, no. 2: 281–318.
Ferrari, S. 2012. A longitudinal study of complexity, accuracy and fluency variation in second language development. In *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*, eds. A. Housen, F. Kuiken and I. Vedder, 277–97. Philadelphia, PA: John Benjamins.
Foster, P., A. Tonkyn and G. Wigglesworth. 2000. Measuring spoken language: a unit for all reasons. *Applied Linguistics* 21, no. 3: 354–75.
Godwin-Jones, R. 2014. Games in language learning: opportunities and challenges. *Language Learning and Technology* 18, no. 2: 9–19.
Halliday, M.A.K. 1985. *Spoken and Written Language*. Melbourne: Deakin University Press.
Horowitz, K.S. 2019. Video games and English as a second language: The effect of massive multiplayer online video games on the willingness to communicate and communicative anxiety of college students in Puerto Rico. *American Journal of Play* 11, no. 3: 379–410.
Housen, A. and F Kuiken. 2009. Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics* 30, no. 4: 461–73.
Hung, H.T., J.C. Yang, G.J. Hwang, H.C. Chu and C.C. Wang. 2018. A scoping review of research on digital game-based language learning. *Computers and Education* 126: 89–104.
Ibrahim, K. 2018. Player-game interaction: an ecological analysis of foreign language gameplay activities. *International Journal of Game-Based Learning* 8, no. 1: 1–18.
Ibrahim, K. 2022. Understanding L2 use in MMOGs: a linguistic framework of digital gaming. *Foreign Language Annals* 55, no. 4: 1169–1187.
Jabbari, N. and Z.R. Eslami. 2019. Second language learning in the context of massively multiplayer online games: a scoping review. *ReCALL* 31, no. 1: 92–113.
Jabbari, N. and Z.R. Eslami. 2023. Negotiations for meaning in the context of a massively multiplayer online role-playing game. *Language Learning & Technology* 27, no. 1: 1–28.
Johansson, V. 2008. Lexical diversity and lexical density in speech and writing: a developmental perspective. *Working Papers in Linguistics* 53: 61–79.
Kyle, K. 2016. Measuring syntactic development in L2 writing: fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication. Doctoral dissertation, Georgia State University, Atlanta, GA.
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, no. 4: 590–619.
Larsen-Freeman, D. 2009. Adjusting expectations: the study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics* 30, no. 4: 579–89.
Larsen-Freeman, D. and L Cameron. 2008. Research methodology on language development from a complex systems perspective. *The Modern Language Journal* 92, no. 2: 200–13.
Laufer, B. and P. Nation. 1995. Vocabulary sise and use: lexical richness in L2 written production. *Applied Linguistics* 16, no. 3: 307–22.
Lave, J. and E. Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
Li, J. 2020. A systematic review of video games for second language acquisition. In *Handbook of Research on Integrating Digital Technology with Literacy Pedagogies*, eds. P.M. Sullivan and J.L. Lantz, 472–99. Hershey, PA: IGI Global.
Li, K., M. Peterson and Q. Wang. 2022. Out-of-school language learning through digital gaming: a case study from an activity theory perspective. *Computer Assisted Language Learning* 35, 1–29.
Lu, X. 2012. The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal* 96, no. 2: 190–208.
MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk*, 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates.
Malvern, D. and B.J. Richards. 2002. Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing* 19, no. 1: 85–104.
Malvern, D., B. Richards, N. Chipere and P. Durán. 2004. *Lexical Diversity and Language Development: Quantification and Assessment*. New York: Palgrave Macmillan.
McKee, G., D. Malvern and B. Richards. 2000. Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing* 15, no. 3: 323–337.

Menard, S.W. 2008. *Handbook of Longitudinal Research: Design, Measurement, and Analysis*. San Diego, CA: Academic Press.
Merriam, S.B. and E.J. Tisdell. 2016. *Qualitative Research: A Guide to Design and Implementation*. San Francisco, CA: John Wiley and Sons.
MMO Populations. n.d. World of Warcraft player count. MMO Populations. Accessed June 1, 2023. https://mmo-population.com/r/wow/.
Morris, L. and T. Cobb. 2004. Vocabulary profiles as predictors of TESL student performance. *System* 32, no. 1: 75–87.
Newgarden, K., D. Zheng and M. Liu. 2015. An eco-dialogical study of second language learners' World of Warcraft (WoW) gameplay. *Language Sciences* 48: 22–41.
Norris, J.M. and L. Ortega. 2009. Towards an organic approach to investigating CAF in instructed SLA: the case of complexity. *Applied Linguistics* 30, no. 4: 555–78.
Ortega, L. 2003. Syntactic complexity measures and their relationship to L2 proficiency: a research synthesis of college-level L2 writing. *Applied Linguistics* 24, no. 4: 492–518.
Peterson, M. 2012a. Learner interaction in a massively multiplayer online role playing game (MMORPG): a sociocultural discourse analysis. *ReCALL* 24, no. 3: 361–80.
Peterson, M. 2012b. Language learner interaction in a massively multiplayer online role-playing game. In *Digital Games in Language Learning and Teaching*, ed. H. Reinders, 70–92. Basingstoke: Palgrave Macmillan.
Peterson, M. 2016. The use of massively multiplayer online role-playing games in CALL: an analysis of research. *Computer Assisted Language Learning* 29, no. 7: 1181–94.
Peterson, M. 2021. Digital simulation games in CALL: a research review. *Computer Assisted Language Learning* 34, 1–24.
Peterson, M. and N. Jabbari. 2022. Digital games and foreign language learning: context and future development. In *Digital Games in Language Learning: Case Studies and Applications*, eds. M. Peterson and N. Jabbari, 1–13. Oxon: Routledge.
Peterson, M., J. White, M.S. Mirzaei and Q. Wang. 2022. A review of research on the application of digital games in foreign language education. In *New Technological Applications for Foreign and Second Language Learning and Teaching*, eds. M. Kruk and M. Peterson, 68–92. Hershey, PA: IGI Global.
Poole, F.J. and J. Clarke-Midura. 2020. A systematic review of digital games in second language learning studies. *International Journal of Game-Based Learning (IJGBL)* 10, no. 3: 1–15.
Rama, P.S., R.W. Black, E. van Es and M. Warschauer. 2012. Affordances for second language learning in World of Warcraft. *ReCALL* 24, no. 3: 322–38.
Rankin, Y., R. Gold and B Gooch. 2006. 3D role-playing games as language learning tools. *EuroGraphics (Education Papers)* 25, no. 3: 33–8.
Rankin, Y., D. Morrison, M. McNeal, B. Gooch and M. W Shute. 2009. Time will tell: In-game social interactions that facilitate second language acquisition. In *Proceedings of the 4th International Conference on Foundations of Digital Games*, ed. R. Young, 161–8. New York: ACM.
Read, J. 2000. *Assessing Vocabulary*. Oxford: Oxford University Press.
Reinders, H. and S Wattana. 2011. Learn English or die: The effects of digital games on interaction and willingness to communicate in a foreign language. *Digital Culture and Education* 3, no. 1: 3–29.
Reinders, H. and S Wattana. 2015a. The effects of digital game play on second language interaction. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)* 5, no. 1: 1–21.
Reinders, H. and S Wattana. 2015b. Affect and willingness to communicate in digital game-based learning. *ReCALL* 27, no. 01: 38–57.
Reinhardt, J. 2019. *Gameful Second and Foreign Language Teaching and Learning*. Basingstoke, England: Palgrave Macmillan.
Reinhardt, J. 2021. Not all MMOs are created equal: a design-informed approach to the study of L2 learning in multiplayer online games. In *Digital Games and Language Learning*, eds. M. Peterson, K. Yamazaki and M. Thomas, 69–87. London: Bloomsbury.
Reinhardt, J. and S.L. Thorne. 2016. Metaphors for digital games and language learning. In *The Routledge Handbook of Language Learning and Technology*, eds. F. Farr and L. Murray, 415–30. Oxford, UK: Routledge.
Reinhardt, J. and S.L. Thorne. 2020. Digital games as language-learning environments. In *Handbook of Game-Based Learning*, eds. J. Plass, R. Mayer and B. Homer, 409–35. Cambridge: Mass.: MIT Press.
Segalowitz, N. 2010. *Cognitive Bases of Second Language Fluency*. New York, NY: Routledge.
Skehan, P. 2003. Task-based instruction. *Language Teaching* 36, no. 1: 1–14.
Skehan, P. 2009. Modelling second language performance: integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics* 30, no. 4: 510–532.
Skehan P., ed. (2014). *Processing Perspectives on Task Performance*. Amsterdam: John Benjamins.
Skehan, P. and P. Foster. 1999. The influence of task structure and processing conditions on narrative retellings. *Language Learning* 49, no. 1: 93–120.
Steinkuehler, C. 2007. Massively multiplayer online gaming as a constellation of literacy practices. *E-Learning and Digital Media* 4, no. 3: 297–318.

Sundqvist, P. 2019. Commercial-off-the-shelf games in the digital wild and L2 learner vocabulary. *Language Learning and Technology* 23, no. 1: 87–113.
Sundqvist, P. and P Wikström. 2015. Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. *System* 51: 65–76.
Swain, M. 2000. The output hypothesis and beyond: mediating acquisition through collaborative dialogue. In *Sociocultural Theory and Second Language Learning*, ed. J. Lantolf, 97–114. Oxford, UK: Oxford University Press.
Sylvén, L.K. and P Sundqvist. 2012. Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL* 24, no. 3: 302–21.
Thorne, S.L. 2008. Transcultural communication in open internet environments and massively multiplayer online games. In *Mediating Discourse Online*, ed. S. Magnan, 305–27. Amsterdam: Benjamins.
Thorne, S., I. Fischer and X. Lu. 2012. The semiotic ecology and linguistic complexity of an online game world. *ReCALL* 24, no. 3: 279–301.
Thorne, S.L. and T. Tasker. 2011. Sociocultural and cultural-historical theories of language development. In *Routledge Handbook of Applied Linguistics*, ed. J. Simpson, 487–500. New York: Routledge.
Tomasello, M. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
Ure, J. 1971. Lexical density: a computational technique and some findings. In Talking about Text, ed. M. Coultard, 27–48. Birmingham: English Language Research, University of Birmingham.
Verspoor, M. and H Behrens. 2011. Dynamic systems theory and a usage-based approach to second language development. In *A Dynamic Approach to Second Language Development: Methods and Techniques*, eds. M. Verspoor, K. de Bot and W. Lowie, 25–38. Amsterdam: John Benjamins.
Wenger, E. 1998. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge, UK: Cambridge University Press.
Yee, N. 2006. The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Presence* 15, no. 3: 309–29.
Yin, R.K. 2018. *Case Study Research and Applications: Design and Methods*. Thousand Oaks, CA: SAGE.
Zheng, D., K. Newgarden and M.F. Young. 2012. Multimodal analysis of language learning in World of Warcraft play: languaging as values-realizing. *ReCALL* 24, no. 3: 339–60.

Appendix

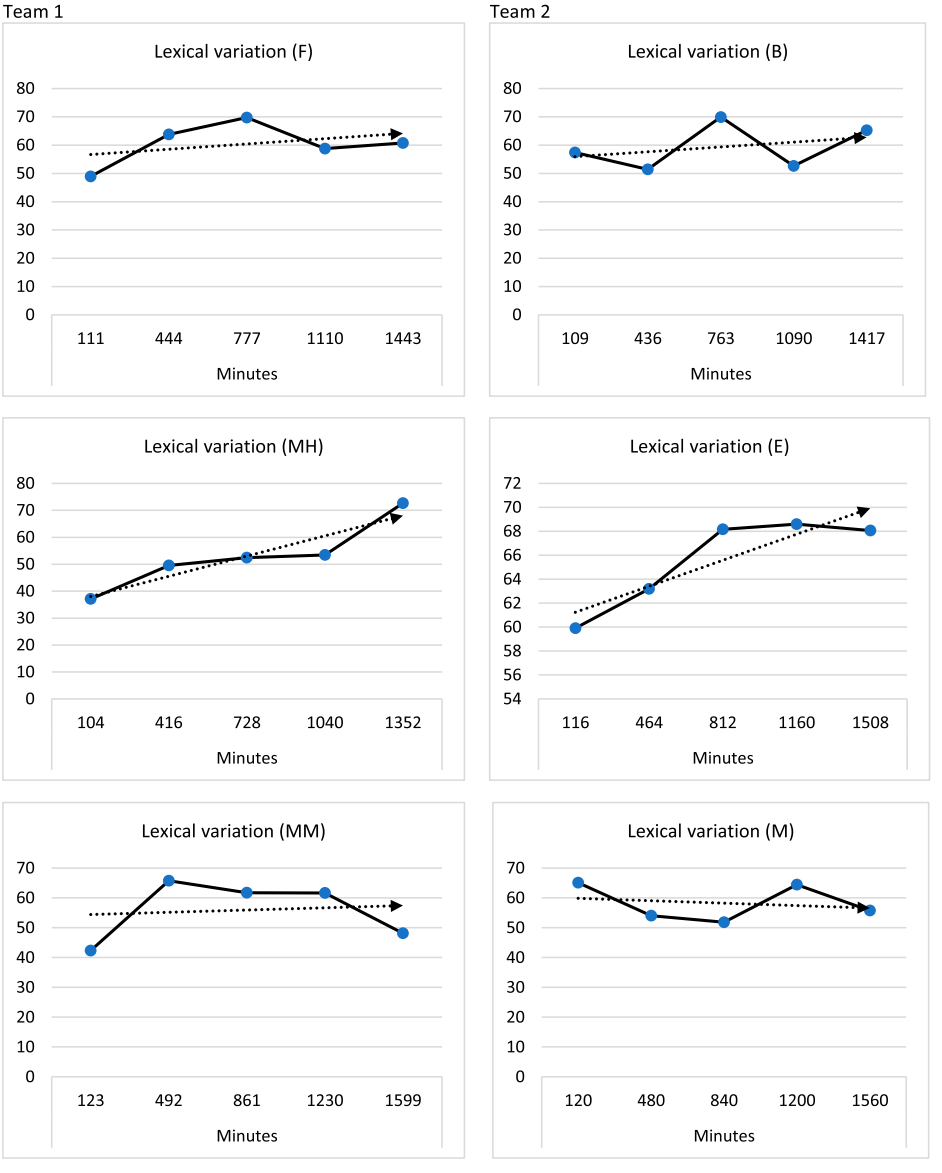

Figure A1. Variations of D values over time.

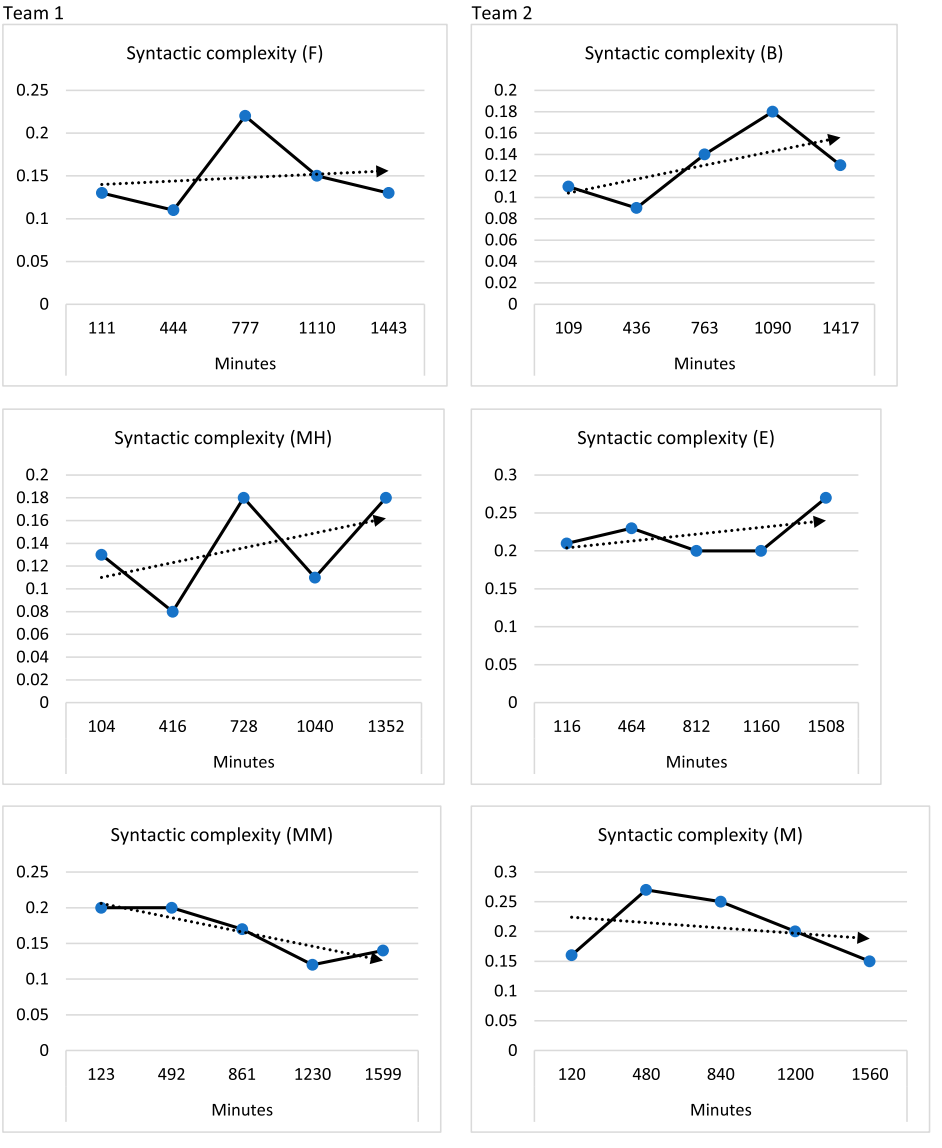

Figure A2. Variations of the average number of subordinate clauses per AS-unit over time.

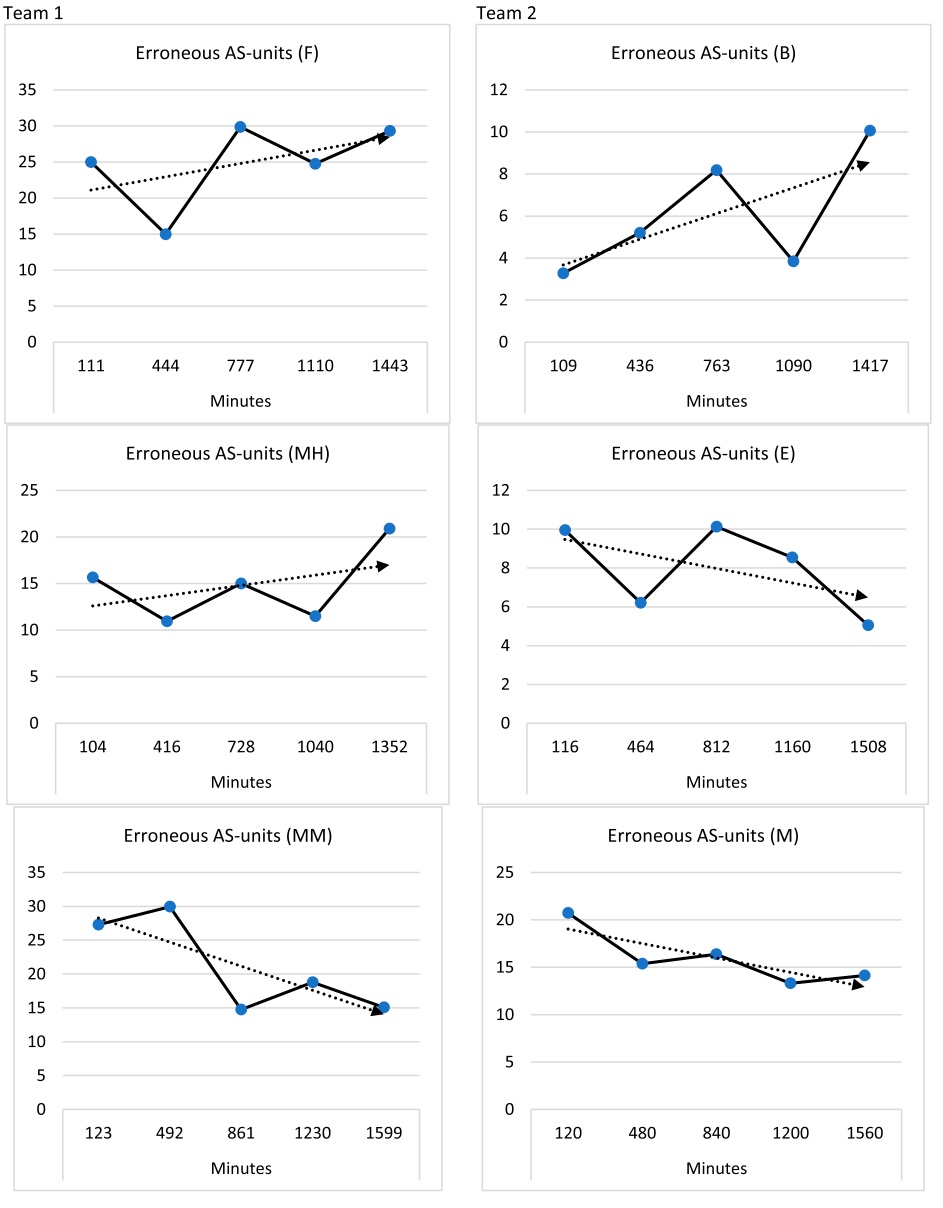

Figure A3. Variations of the % of erroneous AS-units over time.

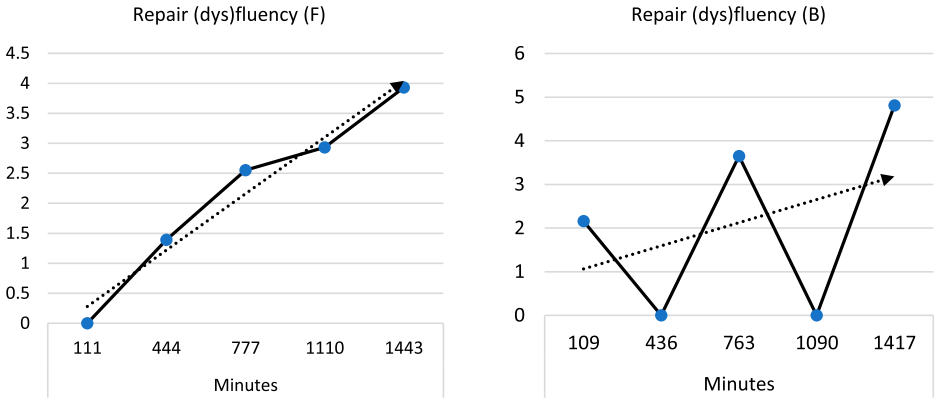

Figure A4. Variations in the average number of (dys)fluency markers per minute (F's and B's data).