10.18686/neet.v2i5.4629

Investigating the Effect of TBLT Approach on Undergraduates' Oral Production Via Ismart Platform

Xinran Wang, Yanfeng Liu, Yanyan Liu

College of Foreign Languages of North China University of Science and Technology, Tangshan 063210, China

Abstract: Under the significant background of online learning, investigating the effectiveness of artificial intelligence (AI) assisted task-based language teaching (TBLT) Approach in enhancing English as a Foreign Language (EFL) oral production is crucial. This study, conducted over 10 weeks, aimed to assess changes in EFL undergraduates' oral production, specifically in terms of complexity, accuracy, and fluency (CAF), following Ismart TBLT instruction. Additionally, the study investigated the evolving relationships among these CAF components, with repeated measures and correlation analyses used to evaluate the data. The findings revealed a significant improvement in the learners' speech rate and the accuracy of their utterances.

Keywords: TBLT; Oral production; Ismart; Complexity; Accuracy; Fluency

1. Introduction

Task-based language teaching (TBLT) has gained prominence as a key approach within communicative language teaching, aiming to create authentic learning environments that enhance real-time communication and develop learners' communicative abilities (Faez & Tavakoli, 2019; Robinson, 2011; Skehan, 1998, 2014). The theoretical underpinnings of TBLT are rooted in the interaction hypothesis, which highlights the significance of negotiating meaning in improving second language (L2) speaking (Long, 1996). Additionally, the output hypothesis (Swain, 2005) is integral to TBLT, as it encourages learners to produce language, identify gaps in their knowledge, and reflect on their progress.

In the current study, technology was integrated with TBLT principles to enhance the teaching and learning experience, focusing on the development of L2 oral language production, defined by the Complexity, Accuracy, and Fluency (CAF) triad. Connecting this approach with AI, the integration of AI tools within the framework of technology-mediated TBLT has the potential to further enhance the language learning experience.

2. Literature Review

With the rapid growth of AI assisted learning and teaching, an increasing amount of research is focused on examining the effectiveness of technology-mediated TBLT on L2 proficiency and production.

2.1 Artificial Intelligence Assisted TBLT and Oral Production

An early longitudinal study by Abrams (2003) explored the development of oral production over a semester by comparing synchronous and asynchronous computer-mediated communication in enhancing oral skills. The study specifically examined the number of words, lexical richness and diversity, and syntactic complexity. Although the synchronous group showed a significant increase in the quantity of language produced, there were no notable differences in lexical richness or syntactic complexity between the groups. However, this study did not address other dimensions of Complexity, Accuracy, and Fluency (CAF), such as fluency and accuracy.However, most of the existing studies on the impact of technology-mediated TBLT on the development of Complexity, Accuracy, and Fluency (CAF) aspects have primarily concentrated on written contexts (Hokamura, 2018).Van der Zwaard and Bannink (2016) examined the negotiation of meaning in a computer-mediated TBLT setting, involving non-native and native English speakers in tasks designed to prompt negotiation.

In conclusion, while existing literature provides valuable insights into the effects of technology-mediated instruction on L2 learning, several limitations are evident. Research focused on blended learning environments introduces variables that may differ significantly from those in purely online settings, potentially impacting language learning outcomes. Furthermore, within technology-mediated TBLT research, there has been a tendency to focus on individual aspects of the CAF triad, rather than adopting a more holistic approach.

2.2 Relationship Among CAF Factors

L2 researchers frequently align the CAF triad with the key components of Levelt's (1989) model of L1 speech production (Robinson,

1997, 2001; Skehan, 1998; Skehan et al., 2016).

Research on CAF often relies on cross-sectional designs, which fail to reveal the evolution of CAF features. Additionally, the long-term relationships among CAF aspects are inconsistently reported, and few studies have explored this in technology-mediated TBLT contexts. This study aims to examine the development of EFL oral production within a technology-mediated TBLT framework and to explore how CAF aspects interact over time.

3. Methodology

3.1 Research Questions

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 Research Method

The study employs a quantitative research method which includes 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.

3.3 Research Setting

3.3.1 Participants

A total of 120 undergraduate students from the Medicine Department at a public university in Hebei participated in this study. Each participant had completed at least 9 years of English education across primary, middle and high schools. Informed consent was obtained from all participants, confirming their understanding of the research ethics and their willingness to participate.

3.3.2 Tasks

Participants completed three tasks at both the beginning and end of the semester as part of the pre-test and post-test procedures. These tasks, adapted from previous research (Albarqi & Tavakoli, 2023), included a personal information task, a football narrative prompt, and a museum narrative prompt. Picture-based narratives and personal information tasks are commonly used in L2 classes and considered ecologically valid in L2 research (Garcia-Ponce & Tavakoli, 2022; Préfontaine & Kormos, 2016; Tavakoli & Foster, 2011).

3.4 Implementation of TBLT Instructions

The study was conducted physically during the September semester as part of a university level project. Participants were enrolled in an College English Course (CEC-1) program, which provided 4 hours of weekly instruction focused on enhancing skills in speaking, writing, reading, and listening, with utmost 1 hour dedicated to each skill. The course used the textbook IEnglish (Wang, 2019) and was delivered by the researcher. The tasks, implemented via the Ismart, were designed according to TBLT principles (Ellis, 2003; Willis & Willis, 2007). Previous L2 research suggests that these phases can affect CAF production.

In the while-task phase, students work in small groups using audio conferencing tools on Ismart. The post-task phase, or language focus phase, involves the instructor identifying errors, providing feedback, and guiding students in the correct use of grammar and vocabulary (Willis, 1996). Unlike traditional approaches, TBLT focuses on meaning, with learners primarily concentrating on content and its semantic and pragmatic aspects. The linguistic form is addressed in the final post-task phase, reflecting a shift towards a more communicative and meaning-oriented pedagogy.

3.5 Data Collection

Data collection contains two phases, pre-test phase and post-test phase. The participants were required to complete three oral tasks at the beginning and end of the semester, with a time interval of 10 weeks between the two phases.

3.5.1 Quantitative Data Collection

The participants were instructed to speak for 1 minute in each task, totalling approximately 360 minutes across the two phases. After transcribing the data, it was segmented into AS-units, following guidelines introduced by Foster et al. (2000). Fluency includes three dimensions: speed, repair, and breakdown (silent/filled pauses). To ensure coding reliability, a second rater coded 10 % of the data for interrater reliability. Disagreements were resolved through discussion between the raters.

4. Results

Descriptive statistics of CAF measures in the two phases are presented in Table 4.1. Most CAF measures indicate an improvement from phase 1 to phase 2 as illustrated in Figure 4.1.

Table 4.1 Descriptive statistics of CAF measures in the two phases

Measures	Phase 1 *M* (SD)	Phase 2 *M* (SD)
MLU	33.67(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	52.1 (20.83)	50.39 (22.82)
Repair	11.26 (6.48)	12.1 (7.41)
SpM	365(97)	403(96)

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

As seen in Table 4.1, the MLU (M=33.67 phase 1 and 35.78 phase 2, SD=6.53 phase 1 and 6.66 phase 2).The silent pauses equals 66.88 in phase 1 and 67.61 in phase 2.The filled pauses and repair of phase 1 lies in 52.1 and 11.26. While the silent pauses and filled pauses in phase 2 are 50.39 and 12.1.

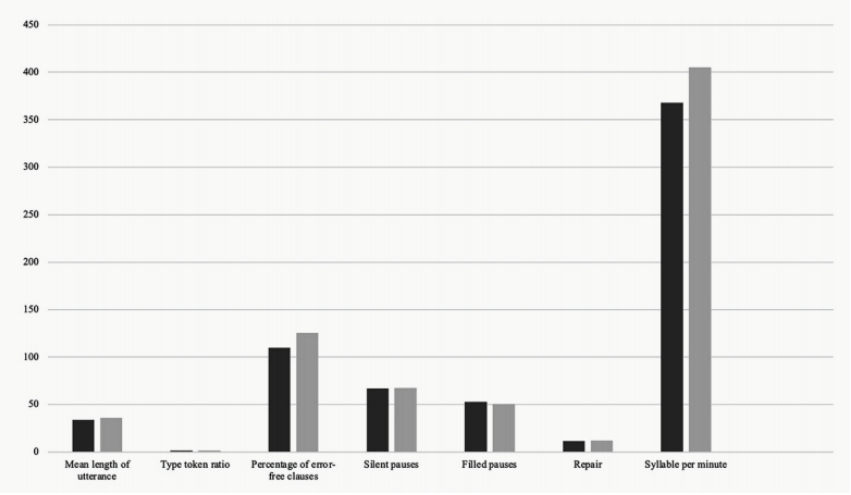

Figure 4.1 Means of CAF measures in the two phases

Figure 4.1 shows a significant change (Wilks' Lambda = 0.624; F = 2.67, p = 0.028; η2 = 0.376), indicating alterations in CAF features following the period of online instruction aligned with TBLT.

Table 4.2 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

*$p < 0.05$.

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.

Table 4.3 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.

As shown in Table 4.3, the result of within-subjects analysis of variance revealed significant differences in two of the measures. Specifically, there was a notable increase in accuracy, measured by 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 indicates that students' oral performance improved significantly in terms of accuracy after 10 weeks. Tables 4.4 and 4.5 reveal two key patterns: the relationships among CAF dimensions and those among fluency features.

Table 4.4 Pearson product-moment correlation: CAF measures in phase

Measures		MLU	TTR	PErFrC	Filled pauses	Silent pauses	Repair	SpM
MLU	Pearson correlation	1	−0.249	−0.011	−0.255	0.377*	0.167	0.415**
			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.

Table 4.4 indicates no significant correlations between accuracy (PErFrC), lexical diversity (TTR), and syntactic complexity (MLU) in the initial phase, suggesting these dimensions were independent at that time. However, fluency features showed significant correlations with other CAF dimensions. Notably, there was a strong positive correlation between syntactic complexity (MLU) and syllables per minute ($r = 0.42$, $p = 0.010$), and syntactic complexity also correlated with silent pauses ($r = 0.38$, $p = 0.019$).

Additionally, there were significant negative correlations between lexical diversity (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 indicates that EFL learners who use a wider range of vocabulary tend to speak more slowly, with fewer silent pauses and self-repairs. Accuracy also showed a negative correlation with filled pauses ($r = -0.35$, $p = 0.034$), suggesting that more accurate speakers pause less frequently. Among fluency features, repair was positively 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 ($r = 0.33$, $p = 0.046$).

Table 4.5 Pearson product-moment correlation: CAF measures in phase 2

		MLU	TTR	PErFrC	Filled pauses	Silent pauses	Repair	SpM
MLU	Pearson correlation	1	−0.021	−0.261	0.138	0.308	0.389*	0.195
			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 4.5 presents the correlations among CAF dimensions during the second phase. Similar to Phase 1, there were no significant correlations between syntactic complexity, lexical diversity, and accuracy, suggesting that these dimensions may develop independently. However, fluency features showed correlations with other dimensions.

Specifically, there was a weak positive correlation between syntactic complexity (MLU) and repair ($r = 0.39$, $p = 0.016$), indicating that more syntactically complex utterances tend to involve more repairs. Additionally, a strong negative correlation was found between lexical diversity (TTR) and repair ($r = -0.56$, $p = 0.000$), suggesting that a broader lexical range in EFL speech corresponds with fewer repairs. This consistent relationship between lexical diversity and repair was observed across both phases.

Accuracy (PErFrC) also demonstrated a significant negative correlation with filled pauses ($r = -0.35$, $p = 0.032$), indicating that students

producing more accurate speech tend to have fewer filled pauses—a trend that remained stable over time. Moreover, in the second phase, accuracy 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 suggests that more accurate speech in the second phase is delivered more quickly, with fewer repairs and filled pauses.

In conclusion, accuracy, lexical diversity, and syntactic complexity appeared to develop independently, while fluency features were interconnected with other CAF dimensions. Consistent relationships across both phases include the link between accuracy and fewer filled pauses, lexical diversity with fewer repairs, and repairs with either filled or silent pauses.

5. Discussion

5.1 Significant Improvements on Accuracy and Speed Fluency

This study revealed significant improvements in accuracy and speed fluency from phase 1 to phase 2. By the end of the semester, EFL learners were speaking faster and with greater accuracy. These findings are consistent with previous L2 studies on oral speech development during study abroad, which have similarly shown increased speech rates after exposure. This trend has been observed both in short stays of three to four weeks and in longer semesters, indicating that similar progress in EFL oral development may occur over a semester of online study.

5.2 Significant Improvements on Accuracy, Lexical Diversity, and Syntactic Complexity

This study revealed that accuracy, lexical diversity, and syntactic complexity developed independently of each other. This suggests that improvements in one aspect did not necessarily coincide with improvements in the others. McManus et al. (2021) observed these relationships over a 21-month period and found that the correlation between accuracy and complexity only became apparent in the final stage of their study, indicating that such connections may require extended time to develop. Similarly, Leonard and Shea (2017) initially identified positive correlations between fluency, lexical complexity, and accuracy in a pre-test phase. Fluency features, in contrast, showed a complex interplay with other CAF aspects across the two phases of the study. In the first phase, there were strong negative correlations between lexical diversity and silent pauses, repair, and syllables per minute.

In conclusion, the accuracy rate showed significant improvement over time. The study found a consistent positive relationship between repair and breakdown features (either filled or silent pauses) across both phases, which suggests that when speakers detect an error in their speech, they pause—using either filled or silent pauses—to plan their subsequent self-correction. Previous research has shown that mid-clause pauses are associated with lower proficiency, while end-clause pauses are linked to higher proficiency levels (Tavakoli 2011). Future research should explore these aspects of EFL fluency and examine the development of speech rate within a technology-mediated TBLT context.

Referenes

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

[2] 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.

[3] 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.

[4] London & New York: Bloomsbury Academic. Bygate, Martin. 2016. Sources, developments and directions of task-based language teaching. Language Learning Journal 44(4). 381–400.

[5] 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.

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

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

[8] 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.

[9] 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.

[10] more developmentally based measures of second language acquisition. Applied Linguistics 35(5). 607–614.

[11] 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.
[12] Levelt, Willem J. M. 1989. Speaking: From intention to articulation, vol. 1. Cambridge: MIT Press.
[13] 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.
[14] 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.
[15] 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.
[16] 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.
[17] Pallotti, Gabriele. 2009. CAF: Defining, refining and differentiating constructs. Applied Linguistics 30(4). 590–601.
[18] Patanasorn, Chomraj. 2010. Effects of procedural content and task repetition on accuracy and fluency in an EFL context. Unpublished PhD dissertation: Northern
[19] 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.
[20] Robinson, Peter. 2001. Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. Applied Linguistics 22(1). 27–57.
[21] Skehan, Peter. 2009. Modeling second language performance: Integrating complexity, accuracy, fluency, and lexis. Applied Linguistics 30. 510–532.
[22] 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.
[23] 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.
[24] Smith, Bryan & Marta González-Lloret. 2021. Technology-meidated task-based language teaching: A research agenda. Language Teaching 54(4). 518–534.
[25] 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.
[26] Rod Ellis (ed.), Planning and task performance in a second language, 11, 239–273. Amsterdam: John Benjamins.
[27] 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.
[28] Westbrook, Carolyn, Lida Baker & Chris Sowton. 2019. Unlock 2: Reading, writing & critical thinking. Cambridge: Cambridge University Press.
[29] Willis, Jane. 1996. A framework for task-based teaching. Harlow & England: Longman.
[30] 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.

Funding: The partially achievement of the project " The Effect of TBLT on College English Course Oral Teaching Via Ismart Platform". (ZL2415)