The effect of digital game-based language learning mobile application on the development of complexity, accuracy, and fluency in foreign language monologic oral production among Chinese learners of English as a foreign language

Feifei Han
University of Sydney

Zehua Wang
Shaanxi Xueqian Normal University

The study reported the effect of a digital game-based language learning (DGBLL) mobile application "Speaking English Fluently – An Automated Scoring Artificial Intelligent Tutoring System on Spoken English" on the complexity, accuracy, and fluency in foreign language (FL) monologic oral production among 31 second year Chinese university learners of English as a foreign language (EFL). The participants' monologic oral production was measured in the first (week 1) and last week (week 21) of a semester using the same narrative picture description task. The oral production was audio-recorded and transcribed. Both the transcripts and audio-files were analyzed on the complexity, accuracy, and fluency dimensions. The complexity was measured using the number of Mean (M) words per T-unit, the accuracy dimension was measured using the number of repairs and errors per 100 words; and the fluency dimension was measured via speech rate (i.e., number of words per minute), and M length of pauses. Students were required to download the mobile application and followed the monological practice section twice a week for 30 minutes each time. Using paired sample t-tests, we found that even though the participants' repair rate and speech rate remained unchanged, they produced more complex monologic speech, had significantly fewer errors, and reduced average length of pauses after 20 weeks treatment using the mobile application, demonstrating a positive effect of the DGBLL mobile application on FL learners' monologic oral production.

Introduction

Technology has undoubtedly opened a new era of human being's experience in every domain, including students' learning experience in higher education. With the advent of new technology in mobile and tablets, hundreds of thousands of mobile applications have been developed to facilitate students' learning. Mobile applications enable students to have more freedom than ever before to make decisions as to when, where, and how to study. Among different kinds of educational mobile applications, game-based learning applications have always been popular, because they are able to foster students' intrinsic motivation, to satisfy students' curiosity, to enhance learners' enjoyment, and to improve students' problem-solving abilities in the process of learning (Dickey, 2011; Gee, 2007; Sung, Hwang, & Yen, 2014). In recent years, the use of digital games for the purpose of learning and teaching in foreign language (known as digital game-based language learning, DGBLL) have been an emergent research area (Cornillie, Thorne, & Desmet, 2012). In a

special issue focusing on DGBLL, Cornille et al. identified five major research themes in the DGBLL, namely (1) theory development, which integrate concepts of digital game-based learning and the theories of FL acquisition and teaching; (2) design theme, which evaluates technological aspects of one or a number of digital games in language learning and teaching; (3) pedagogical theme, which is concerned with teachers' self-reflection and evaluation of effectiveness of using certain digital games in language teaching; (4) experimental studies, which compare the effect of DGBLL intervention with a control group of traditional intervention; and (5) non-experimental empirical studies, which report the effects of DGBLL on development of one skill in FL learning. The research we reported is in the last theme, in which we examined the effect of using of a DGBLL mobile application for 20 weeks on the development of Chinese EFL learners' monologic oral production based on the complexity, accuracy, and fluency dimensions.

The effect of DGBLL on FL learning

An increasing number of researchers and educators have recognized how technology can be pedagogically exploited to facilitate acquisition and engagement in FL learning in recent years (Shadiev, Hwang & Huang, 2017; Ushioda, 2013). In FL classrooms, more and more teachers have also employed various digital games and applications. The advantages of using these digital games and applications has been described as interactive, enjoyable, exciting, stimulating, and well-structured (Vandercruysse, Vandewaetere, & Clarebout, 2012), and the game-based experience is said to be transferable into language learning (Chiu, Kao, & Reynolds, 2012). Past research has been carried out to examine the gaming environment on FL learning and suggests that when the primary focus of the gaming environment is not on linguistic aspects, the digital games may not necessarily be conducive to language learning (e.g., deHaan, Reed, Kuwada, 2010), whereas gaming environment which supplies sufficient opportunities of repetition of linguistic elements tends to lead to positive gains of language acquisition (e.g., Zheng, Young, Wagner, & Brewer, 2009). Research in DGBLL has also shown that different types of digital games have differential effects on language learning. In a meta-analysis, Chiu et al. (2012) demonstrates that drill and practice digital games only yield small positive effect (0.41 to 0.44) whereas meaningful and engaging digital games have large positive effect size (0.84 to 1.11). Studies in DGBLL have also targeted on the development of different language skills, in particular, there are more studies focusing on vocabulary learning (e.g., Cobb & Horst, 2011; deHaan et al., 2010; Yip & Kwan, 2006). There is a dearth of studies which examine the digital games on the development of FL oral production.

Measurement of FL production

In the literature of FL learning, it is widely acknowledged that learner production is multidimensional in nature, hence, it is difficult to capture such production using a single measurement (Housen & Kuiken, 2009; Norris & Ortega, 2009; Pallotti, 2009). Attempting to represent FL production comprehensively, a number of frameworks have been constructed and testified. For instance, established and developed by the Council of Europe (2011), Common European Framework of Reference for Languages (CEFRL), is a guideline which describes FL learners' performance using communicative competence construct. The framework states in details what constitutes different levels of competence and performance in the four skills (i.e., reading, writing, listening, and speaking) of FL learning.

From a componential perspective, learner performance and production in FL can also be measured by multiple components, namely complexity, accuracy, and fluency, which are known as CAF triad (Skehan, 1998). The three dimensions are widely adopted to describe and assess learner competence and performance, in particular in writing and speaking domains (Ellis & Barkhuizen, 2005; Housen & Kuiken, 2009; Housen, Kuiken, & Vedder, 2012; Ortega, 2003; Skehan, 2009; Yuan & Ellis, 2003). Complexity represents the breadth and depth of the language production (Ellis, 2003); accuracy describes the level of conformity of language production to certain norms (Pallotti, 2009); and fluency indicates the extent of automaticity of language production (Wolfe-Quintero, Inagaki, & Kim, 1998). According to different operationalization, each dimension can be gauged by using different indices (Rosmawati, 2014). In our study, we employed the CAF triad to investigate development of Chinese EFL learners' monologic oral production after using a DGBLL mobile application for 20 weeks.

Method

Participants

The participants were 31 second-year students, who were enrolled in a four-year Bachelor degree in English education in early childhood at a national university in China. All the participants were females as this major tends to attract female students. Their ages were between 19 and 21 with a M of 20. All the students reported that they had learnt English as a FL for approximately 10 years.

Instruments

Monologic narrative picture description task for pre- and post- oral production. We used a monologic narrative picture description task to measure the participants' oral production for two reasons. First, monologic tasks are not influenced by interactional factors as in the dialogical tasks. Second, we selected the narrative genre because the participants had intensive practice of narratives in English oral production than the other genre as indicated by their English teachers. We used the same picture description task in pre- and post-test because this reduced the variation of task difficulty to the minimal level. Due to the long period between pre- and post-test (20 weeks), the effect of task repetition would also be negligible. The task required the participants to describe four pictures by telling a story in English. The pictures depicted a story about a little girl who accidentally fell into a river and how she was rescued by her dog. The picture description task was piloted with 5 students with the similar background and English proficiency. None of the students had difficulties in understanding and describing the pictures in English, suggesting that the instrument was appropriate to elicit students' English oral production.

DGBLL mobile application. We used "Speaking English Fluently – An Automated Scoring Artificial Intelligent Tutoring System on Spoken English" mobile application to develop students' oral production. The application ranked as one of the most popular mobile applications in

practicing spoken English in China with more than 420 million users. The application was specifically developed for Chinese EFL learners and its special speech recognition system has been created to recognize English speech of Chinese EFL learners. The application allows users to practice spoken English via mobile phone microphones and is able to record learners' speech, which is then analyzed using algorithms integrated with the application in order to compute a score and provide immediate feedback for learners' oral production. The gaming environment of the application is able to engage the learners. First, the application is able to stimulate learners' interests and passion to practice spoken English through game-based elements, such as rewarding users with gold coins and stars, using different ranking systems to give ranks of different users to simulate competitions among game players, and allowing users to break through into different stages like those in other digital games. Second, the topics are diverse so that they can satisfy various needs of different users; and third, the automated scoring system is able to provide immediate feedback and the application can also offer individualized dashboard, which is able to sustain learners' motivation. The students were asked to follow "Imitation of English Monologues" section twice a week for 30 minutes each time to practice monologic oral English.

Procedure

The study had a pre- and a post-test session in the first and last week of a semester with 20-week in between. At the beginning of the semester (week 1), the researchers explained to the students the purposes of the study and invited them to voluntarily participate the study. Then the pre-test, which asked the learners to describe four pictures in a narrative, was conducted to elicit the initial level of monologic oral production of the students. The picture description task was taken away after testing to ensure that students would not have opportunities to practice it between pre- and post-tests. Upon completion of the pre-test, the students were instructed on how to install the mobile application and on what they need to do. During the 20 weeks, the students practiced using their own time and were required to keep their practice recorded on their mobiles each time. After practice, they were also required to keep a log to write down the topics and the duration they practiced. Both the recordings of the practice and the logs were regularly checked by their English teachers to ensure that they followed the instruction. After 20-week practice, the participants' English monologic oral production was examined again by asking them to narrate the same pictures used in the pre-test. Students' narration was audio-recorded.

Coding and data analysis
The audio-recordings were transcribed and both the transcripts and the audio-files were coded in terms of the three dimensions, namely complexity, accuracy, and fluency dimensions (see Table 1). To examine the

complexity dimension, we calculated the Mean (*M*) words per T-unit using an online calculator – Sentence Extractor (http://www.lextutor.ca/tools/ex_sent/). A T-unit is an independent clause and any dependent (subordinate) clauses or non-clausal structures that are attached to or embedded within it (Lennon, 1990). The more the number of *M* words indicated the more complex the oral production was.

To examine the accuracy dimension, we used two indices – number of repairs per 100 words and number of errors per 100 words. The number of repairs per 100 words was calculated by dividing the total number of repairs by the total number of words in the speech and then multiplying 100. According to Foster and Skehan (1996), there are five types of repair, including reformulation, replacement, repetition, false start, and hesitation. Reformulation is defined as repeating a phrase or a clause by modifying any of morphological, syntactical, or word order of the phrase or the clause (e.g., The girl **see**…the girl **saw** a ball). Replacement refers to substituting a phrase or a clause with another phrase or a clause (e.g., The girl is **following**…**chasing** the dog). Repetition is restating exactly the same phrase or clause without any modification (e.g., The girl…**The girl** is angry). False start means completely giving up a phrase or a clause (e.g., **The girl is**…**Her parents** are waving to her). Hesitation is referred to as repetition of a phoneme or a syllable of a word (e.g., The dog **ba**… barked to her parents).

Similarly, the number of errors per 100 words was calculated by dividing the total number of errors by the total number of words in the speech and multiplying 100. We included both grammatical (e.g., Her parents is (are) working on the farm) and lexical errors (e.g., The girl is **catching** (picking) the flower). The fewer the number of repairs per 100 words and the number of errors per 100 words means the more accurate the speech was.

To examine the fluency dimension, we used two indices: speech rate and the *M* length of pauses, which were coded under the assistance of the software Cool Edit Professional 2.0 of the recordings. Speech was expressed in terms of number of words per minute and was calculated by the total number of words divided by the speech length (in minutes). The higher the value of the speech rate represented the more fluent the speech was. The *M* length of pauses was expressed in seconds and was calculated by averaging the length of all the pauses in a speech. The longer the *M* length of pauses indicated the less fluent the speech was. In our study, a pause was identified as a break of 1 second or longer either within a sentence or between sentences. We entered the coded data into SPSS 22 and conducted paired sample *t*-tests to examine if there were significant differences on the five indices of the participants' monologic oral production in English between pre- and post-test.

Table 1. Five indices of the three dimensions of monologic oral production in English

Dimensions	Indices
complexity	*M* words per T-unit
accuracy	repairs per 100 words
	errors per 100 words
fluency	speech rate
	M length of pauses

Results and discussion

Table 2 presents the descriptive statistics of the pre- and post-test. Among these indices, we can see that students have large variation in terms of speech rate as shown in large *SDs* in both pre- and post-test. This index might not only reflect students' fluency in English monologic oral production, but might also be caused by the individual differences in speech rate. We conducted paired sample *t*-tests to examine if the participants' monologic oral production differed between pre- and post-test.

Table 2. Descriptive statistics of pre- and post-test

Dimensions	Indices	Pre-test		Post-test	
		M	*SD*	*M*	*SD*
complexity	*M* words per T-unit	11.94	2.91	15.57	4.05
accuracy	repairs per 100 words	6.07	3.73	5.46	2.30
	errors per 100 words	8.08	2.80	5.99	2.68
fluency	speech rate (per minute)	46.86	14.91	61.33	15.37
	M length of pauses (in seconds)	3.69	2.06	3.10	2.10

In terms of the complexity measure, the paired sample *t*-test showed that the *M* words per T-unit in the post-test (*M* = 15.57, *SD* = 4.05) was significantly more than that in the pre-test (*M* = 11.94, *SD* = 2.91), *t* = -5.1, *p* < .01, indicating that the participants' monologic oral production in English was more complex and they were able to produce longer English sentences after 20-week practice with the mobile application. With regard to the two indices of the accuracy, we found that while the number of repairs per 100 words remained unchanged between pre- (*M* = 6.07, *SD* = 3.73) and post-test performance (*M* = 5.46, *SD* = 2.30), *t* = 0.98, *p* = .33; the participants had significantly fewer errors in the post-test production (*M* = 5.99, *SD* = 2.68) than in the pre-test (*M* = 8.08, *SD* = 2.80), *t* = 3.99, *p* < .01. The significant reduction of the error rate, both grammatical and lexical, may be influenced by the improvement of students' grammatical knowledge and expansion of vocabulary knowledge. The non-change of the repair rate seemed to indicate that the knowledge learnt explicitly (i.e., declarative knowledge) had not been proceduralized (i.e., procedural knowledge), therefore, when the knowledge was used in an online processing task, as in our monologic oral production task, the learners' repair rate did not change.

Similarly, for the two indices of the fluency dimension, we observed that while students' speech rate did not change significantly between pre- (*M* = 46.86, *SD* = 14.91) and post- oral production (*M* = 61.33, *SD* = 15.37), *t* = 1.15, *p* = .25; the *M* length of pauses reduced significantly from the pre-test (*M* = 3.69, *SD* = 2.06) to the post-test (*M* = 3.10, *SD* = 2.10), *t* = -4.95, *p* < .01. In general, our study showed a positive effect of the DGBLL mobile application on the development of Chinese EFL learners' English monologic oral production on all the three dimensions. The mobile application not only enabled students to have much freedom as to when and where to practice spoken English with their mobiles in their free time, but was also able to provide immediate feedback using an automated scoring system. To extend the present study, we will compare the effects of traditional methods of practicing spoken English with the DGBLL mobile applications on FL learners' oral production. In future studies, we will also incorporate multiple indicators of the complexity dimension, such as syntactic complexity and lexical complexity.

References

Cobb, T., & Horst, M. (2011). Does Word Coach coach words? *CALICO Journal*, 28(3), 639-66. https://doi.org/10.11139/cj.28.3.639-661

Cornillie, F., Thorne, S. L., & Desmet, P. (2012). ReCALL special issue: Digital games for language learning: challenges and opportunities. *ReCALL*, 24(03), 243-256. https://doi.org/10.1017/S0958344012000134

Council of Europe. (2011). *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Strasbourg: Council of Europe.

deHaan, J., Reed, W. M., & Kuwada, K. (2010). The effect of interactivity with a music video game on second language vocabulary recall. *Language Learning & Technology*, 14(2), 74-94.

Dickey, M. D. (2011). Murder on Grimm Isle: the impact of game narrative design in an educational game-based learning environment. *British Journal of Educational Technology*, 42(3), 456-469. https://doi.org/10.1111/j.1467-8535.2009.01032.x

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

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*,18(03), 299-323. https://doi.org/10.1017/S0272263100015047

Gee, J.P. (2007). *What video games have to teach us about learning and literacy (2nd ed). New York*: Palgrave Macmillan.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461-473. https://doi.org/10.1093/applin/amp048

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy, and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 1-20). Amsterdam/Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.32.01hou

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40(3), 387-417. https://doi.org/10.1111/j.1467-1770.1990.tb00669.x

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555-578. https://doi.org/10.1093/applin/amp044

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college‑level L2 writing. *Applied Linguistics*, 24(4), 492-518. https://doi.org/10.1093/applin/24.4.492

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30(4), 590-601. https://doi.org/10.1093/applin/amp045

Rosmawati, R. (2014). Dynamic development of complexity and accuracy. Australian Review of *Applied Linguistics*, 37(2), 75-100. https://doi.org/10.1075/aral.37.2.01ros

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press. https://doi.org/10.1177/003368829802900209

Skehan, P. (2009). Modeling second language development: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510-532. https://doi.org/10.1093/applin/amp047

Sung, H. Y., Hwang, G. J., & Yen, Y. F. (2015). Development of a contextual decision-making game for improving students' learning performance in a health education course. *Computers & Education*, 82(1), 179-190. https://doi.org/10.1016/j.compedu.2014.11.012

Vandercruysse, S., Vandewaetere, M., & Clarebout, G. (2012) Game-based learning: A review on the effectiveness of educational games. In Cruz-Cunha, M. M. (Ed.), *Handbook of research on serious games as educational, business, and research tools* (pp. 628-647). Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-4666-0149-9.ch032

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Manoa, HI: Second Language Teaching and Curriculum Center, University of Hawai'i.

Yip, F. W. M., & Kwan, A. C. M. (2006). Online vocabulary games as a tool for teaching and learning English vocabulary. *Educational Media International*, 43, 233-249. https://doi.org/10.1080/09523980600641445

Zheng, D., Young, M. F., Brewer, R. A., & Wagner, M. (2009). Attitude and self-efficacy change: English language learning in virtual worlds. *CALICO Journal*, 27(1), 205-231. https://doi.org/10.11139/cj.27.1.205-231

Note: All published papers are refereed, having undergone a double-blind peer-review process.