

# The Effectiveness of Mobile-Assisted Language Learning (MALL) Applications on the Spoken English Assessments in China's Universities

LI Xiaomeng

University of South Australia (UniSA), Adelaide, Australia

The increased ownership of mobile phone users and the advancement of mobile applications enlarge the practicality and popularity of use for learning purposes among Chinese university students. However, even if innovative functions of these applications are increasingly reported in relevant research in the education field, little research has been in the application of spoken English language. This paper examined the effect of using a Mobile-Assisted Language Learning (MALL) application "IELTS Liulishuo" (speaking English fluently in the IELTS test) as a unit of analysis to improve the English-speaking production of university students in China. The measurement of this mobile application in its effectiveness of validity and reliability is through the use of seven dimensional criteria. Although some technical and pedagogical issues challenge adoptions of MALL in some less-developed regions in China, the study showed positive effects of using a MALL oral English assessment application characterised with Automatic Speech Recognition (ASR) system on the improvement of complexity, accuracy, and fluency of English learners in China's colleges.

*Keywords:* Mobile-Assisted Language Learning (MALL), Artificial Intelligence (AI) education, Automatic Speech Recognition (ASR), China, university students

## Introduction

China has a talent-oriented education tradition, and the expectation for young people is that they can improve their abilities by participating in fierce competition, especially in oral English (Liu, Yang, & Xu, 2019). It attracts instructors' interest in finding out new teaching methods to aid them to achieve this educational purpose. Some innovative learning theories have been a trend in western countries (Liu et al., 2019). In the meantime, there are successful examples in university English for a special purpose in corpus vocabulary with the adoption of mobile apps; and it motivates teachers and learners to shift the focus to Mobile-Assisted Language Learning (MALL). MALL is famous for technical affordances in abundant learning experiences and opportunities for students (Dai & Wu, 2021) at outside class contexts towards active engagements in relevant activities (Zou, Yan, & Li, 2020). An apparent example to illustrate the advantages of MALL is Liulishuo, which is a spoken English production assessment tool. It is assumed that MALL can address Chinese college students' difficulties in IELTS speaking via a unit of analysis of the Liulishuo mobile app. However, little

---

LI Xiaomeng, Master of Education (TESOL), UniSA Education Future, University of South Australia (UniSA), Adelaide, Australia.

research exists on MALL for second language acquisition in China, especially in spoken English. Targeting to fill in the gap in MALL's investigations on oral English assessment in China, this thesis adopts a review of the pros and cons of MALL. This paper aims to address a question in the effectiveness of English oral assessments in MALL to improve Chinese university students' English learning proficiency. Besides, this investigation could shed light on future research in the field of emerging educational technologies in Chinese people's better English learning outcomes.

### **The Benefits of MALL—IELTS Liulishuo as an Example**

MALL is a subset of mobile learning and computer-assisted language learning (Deloitte, 2019). It applies digital phones as tools in virtual classrooms to assist students' language learning. MALL has many benefits for English education. Firstly, MALL prompts target-language (L2) use *procedualisation*. Secondly, mobile phones extend full use of learners' filler-time. It is appreciated that flexible access and convenience of digital phones without time and space limitations create a self-paced and time-efficient virtual learning environment for learners' revision tasks (Sun & Gao, 2020). Finally, mobile phones offer a manageable data-sorting system. Indeed, downloaded speaking learning materials are well-organised in orders of data or other categorised labels (Zou et al., 2020).

In recent years, MALL assessments in AI apps have attracted much attention from Chinese educators and governments. The Automatic Speech Recognition (ASR) function in Artificial Intelligence (AI) has potentials in autonomous learning and independent sharing. A working definition of AI is that it "refers to the reasoning, interacting and learning functions associated with human beings" (Rai, Constantinides, & Sarker, cited in Fu, Gu, & Yang, 2020, p.1677). Indeed, fruitful learning resources in mobile apps enrich learners' obtainable materials outside classrooms (Liu et al., 2019; Dai & Wu, 2021; Zou et al., 2020). For example, in Liulishuo, if it included teachers' guidance in spoken English learning, learner autonomy would be improved.

Additionally, independent sharing in Liulishuo is assumed to be accomplished by learners' sharing daily practice results in their online learning communities (Liu et al., 2019).

The success of the designed assessment is its effectiveness in addressing concerns within seven types of assessments. To prove effectiveness, Liulishuo solves issues that exist in seven characteristics (i.e., formative, objective, direct, performance-referenced, integrated, score reliability, and construct validity) of identified assessment activities.

#### **Formative Assessment**

The main feature of formative assessment is that it provides feedback to learners as a "process assessment", which can guide instructors for subsequent teaching and learning tasks (Carr, 2011). In terms of content validity of IELTS Liulishuo, it meets standardised criteria by IELTS (International English Language Testing System), relevant topic corpus found in various sequencing exercises, such as "building & engineering" (Wang et al., 2020). Moreover, Brown (2004) points out the requirements of formative feedback to achieve learning efficiency. It is believed that several of them ("self-assessment performance" and "after-class teachers' comments") are reflected in learners' self-reflective learning diaries and teachers' synchronised future improvements to track learners' progress in the app (Figure 1). Hence, valid topic coverage and teachers' suggestions indicate the usefulness of this type of assessment.



Figure 1. The assessment report for speaking fluency: Auxiliary rhetorical devices.

### Objective Assessment

According to Carr (2011), responsive questions (e.g., matching) as assessments in objective tests constitute the biggest component. Due to the objective characteristic of the IELTS test, matching learning tasks in the app are assumed to be identical with the objective assessment. Additionally, banking quantitative IELTS speaking practice results are calculated by the app. Based on deep learning, it is elaborated as “intelligent correcting”, which supplies reasonable and synchronous feedback to learners (Deloitte, 2019). Finally, it is reinforced by the app’s numerous and sophisticated datasets collection and data filtering to ensure the reliability of grading. Therefore, eradication of subjectivity is guaranteed in the app’s quantitative data analysis.

### Direct Assessment

In English language education, four macro skills (i.e., listening, speaking, reading, and writing) are assessed for examinees’ English proficiencies in tests directly (e.g., writing something in writing tests) (Carr, 2011). The Liulishuo oral production assessment adopts explicit feedback and grammar-translation approaches to evaluate learners’ speaking skills through relevant activities. For instance, in Extract 1 of the sentence-translation task, it is recognised that learners’ *negotiation of meaning* is obtained through teachers’ mother tongue to clarify teaching demands (Figure 2). Compared to modeling answers, it eventually enhances learners’ self-reflection skills by enriched language knowledge and intends not to make the same mistake twice in future practices. Hence, this assessment fills gaps between students’ current learning levels and intended learning outcomes directly.

### Performance-Referenced Assessment

During the learners’ second language acquisition process for proficiency, language accuracy and closed relationships to the real world are two major evaluations in the *strong sense of language performance assessment* of learning tasks (Carr, 2011). For example, technologies of *computer-assisted pronunciation training*, as well as *Automatic Speech Recognition (ASR)* as the core features of Liulishuo, could prompt Chinese test-takers’ *accent intelligibility*, leading to their fluency in the IELTS examination (Wang et al., 2020).

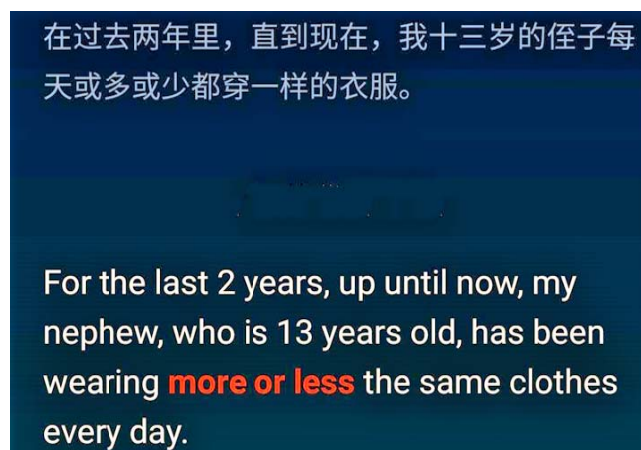


Figure 2. Extract 1. "Sentence-translating: Chinese to English" IELTS speaking task (IELTS Liulishuo, 2020).

### Integrated Assessment

Integrated assessment examines students' multiple abilities in language-in-use tasks (e.g., contextualised tasks) (Carr, 2011). Nevertheless, this scholar raises concerns about three challenges affecting assessment results, which are rubric scoring standards, ineffective interpretations, and limited topic coverage (Carr, 2011). Fortunately, within an AI-guided standardised and personalised assessment, the scientific conceptual framework lies in big data and learning analytics are very valid, which can solve the first issue (Deloitte, 2019). The second issue is addressed by the app's intelligent tutors. Learners are individually allocated with a private intelligent "agent", who plays a similar role as teachers in physical classrooms to clarify any follow-up questions or misunderstandings. Thirdly, a variety of regularly updated topic practices similar to real IELTS speaking tests extend the breadth of covered samples.

### Scorer Reliability

Reliable scores in this assessment depend on qualitative statistics and their consistency. To ensure data and results' reliability, a potential issue needs to be avoided relating to accurate error-measurement, which is affected by the rationale of tests, test-taker's willingness in accomplishments, and the effectiveness of grading tools (Carr, 2011). The role of reliability is its responsiveness of all assessed tasks to the four-criteria marking syllabus in the IELTS test. The rest of the two variables are addressed by Liulishuo's experimental five-round data filtering procedures to produce gold-standard results as sample answers (Figure 3) (Wang et al., 2020). Thus, categorising and discarding unreliable statistics through reviewers' recognitions and group discussions can generate convincing educational evidence being disseminated in the app's grading tasks.

### Construct Validity of the Assessment

Construct validity is specified as the compatibility of its pre-determined task levels with learners' current levels to assign scaffolding tasks to them (Carr, 2011). Indeed, joint evidence proves a so-called term *task difficulty* as a psycholinguistic variable reflected in sequencing tasks that is closely correlated with construct validity (Fulcher, 2003; Ahmadi & Nazari, 2014). Fulcher (2003) further argues *task complexity* contributes to *task difficulty*. Hence, the effectiveness of construct validity seems difficult to achieve from grading tasks. However, as learners become familiar with the ladder-climbing learning process in Liulishuo, they will be successful in connecting different fields of knowledge by activating their prior knowledge to form *declarative knowledge* through timely revisions. In this way, the app can be seen as valid in its content knowledge, and

these variances are easily observed in each procedure of activities. Similarly, breaking down one input flood activity (i.e., topic-relevant vocabularies or sentence-repetition) into several mini-steps with audio-visual modes to suit different learning styles can decrease test-taker's cognitive load. Therefore, it eliminates *task complexity*. Eventually, more engaged learners are able to develop their 21st-century skills (e.g., creativity and problem-solving), and perform well in IELTS tests.

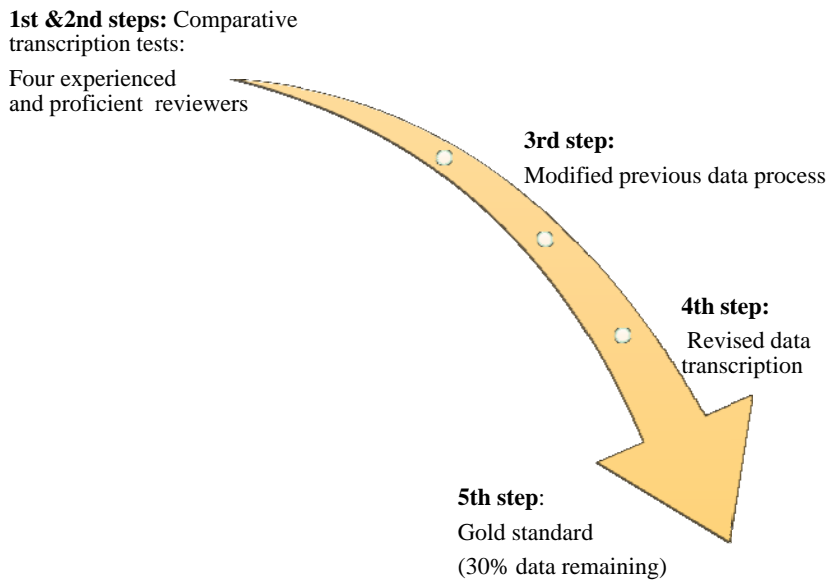


Figure 3. Automatic scoring system in “IELTS Liulishuo” of transcribed data.

### Critique of ASR-Enabled Apps in MALL

Even if multi-dimensional sources and data exist indicating credibility and validity standards of Liulishuo or even if ASR informed MALL in oral English, they would be criticised for discrepancies in pedagogy, learners' resistance, and poor app performances.

#### Pedagogical Issues

One tension caused by ASR-enabled apps is indirect feedback. As Dai and Wu (2021) claim, accuracy-error-correcting largely relies upon learners' self-monitoring, and these apps supply exploitative feedback. Contrasting with informal feedback, Chinese students are accustomed to receiving formal-written feedback from teachers. Furthermore, external social interactions and peer feedback seem to be more effective scaffolding strategies, whereas peer collaboration is rare in these apps. Consequently, MALL may not provide new quality teacher and student interactions due to the disruptive pedagogy (Zou et al., 2020).

#### Sociocultural Issues

Generally, traditional cultural beliefs and a familiar learning atmosphere lead to learners' regional gaps in a reluctance of MALL. One reason is China's pre-determined learning atmosphere. Traditionally, face-to-face classrooms adopting physical books were the most common teaching mode until now in many underdeveloped regions of China (Zou et al., 2020).

Another reason why traditional learners hold a negative view of MALL is the obvious gap between provinces. It has been demonstrated that in developing cities of mid-western provinces of China, university students possess limited familiarity with ASR apps resulting in little ease of use of mobile learning and low-level studying intentions (Sun & Gao, 2020). This is further compensated by learners' insufficient intrinsic motivations in achieving targeted education purposes such as lack of interest in English (Sun & Gao, 2020). Therefore, to some extent, there is a possible issue of learners' resistance in adopting ASR apps or Liulishuo in continuous oral English learning.

### **Technical Issues**

ASR-initiated apps' poor performances in software and hardware lead to doubts of effectiveness as well. Software shortcomings are the shortage of multifunctionality in ASR apps and their built-in web browsers. In other words, ineffective interactions might arise when there are few available ASR apps (Zou et al., 2020). For instance, the number of these apps on iPhones seems more than those available in Android's app store. Similarly, Apple's Safari browser is better than Android phone internet browsers because it has plug-in flash players to play downloaded materials from ASR apps. Hardware insufficiencies include tiny screens and transient computational batteries (Zou et al., 2020). This implies learning outcomes are affected in dictating tasks that require learners to type words with keyboards.

To conclude, tensions and challenges in considerations of pedagogy, socio-cultural elements, and techniques in digital phones need to be taken into account for ASR-enabled apps or Liulishuo's learning efficiencies.

### **Corresponding Solutions to Existing Issues in MALL**

From the pedagogy's perspective, identified level-appropriate teaching content and self-regulated learning (SRL) self-studying strategies are suggested to adopt in MALL. Firstly, sustainable education purposes and a stress-free atmosphere are pre-requisites of learners' intrinsic motivation (Dai & Wu, 2021). Hence, it implies that instructors figure out "known" and "unknown" scaffolded learning content to suit differentiated learners' current speaking levels and study speed. For example, based on placement tests of ASR-enabled apps, teachers could select ability-competent materials to aid learners in deciding short-term speaking goals and making them easier to accomplish anticipated outcomes. Secondly, once learners are proficient in SRL, their self-monitoring abilities and learner autonomy will be enriched. Indeed, Tsai (cited in David & Wu, 2021) proposed the three-step "sandwich model" could orchestrate peer feedback and ASR feedback in a harmonised manner. It is described as: mediating ASR technologies to inform self-practice; interacting with peers via networking online platforms; using ASR apps again for self-checking; and coming up with additional practices to deal with unresolved issues (Dai & Wu, 2021) (Figure 4). In this way, in Liulishuo, learners' recognitions of "noticing" mistakes or confusions are raised by noting down self-reflections in a diary (Table 1). As a result, learners gain self-assessment abilities, which are critical in autonomous learning developments.

From the app design's perspective, it is important for tool developers to adjust apps' design properly to meet each learners' needs and promote the role of MALL for infrequent users. It is true that only if ASR apps implement needs analysis of students (Liu et al., 2019; Dai & Wu, 2021; Zou et al., 2020; Sun & Gao, 2020; Fu et al., 2020; Carr, 2011), the learning environment's preferences of them would be changed to mobile phones, which could also result in enhanced behavior intentions to use the apps consistently (Sun & Gao, 2020). To

reduce gaps between app designers and teachers in instructional teaching methods in virtual classrooms, ASR apps must incorporate enjoyable materials, versatile functions, and user-review-based app-development plans (Sun & Gao, 2020). In addition, to increase traditional learners' familiarity with MALL apps in regional provinces of China, policymakers and other stakeholders are encouraged to publicise the advantages of intelligent education until it becomes a trend (Liu et al., 201). In the meanwhile, there should be reconsideration of cost-effective artifacts for users to subscribe to apps' memberships also contributing to satisfactory learning processes (Liu et al., 201). Typical examples are that after the promotion of top-rated ASR apps via social media, app-designers can advertise point-accumulating purchasing systems for multi-artifacts or large amounts of discounts at holiday breaks.

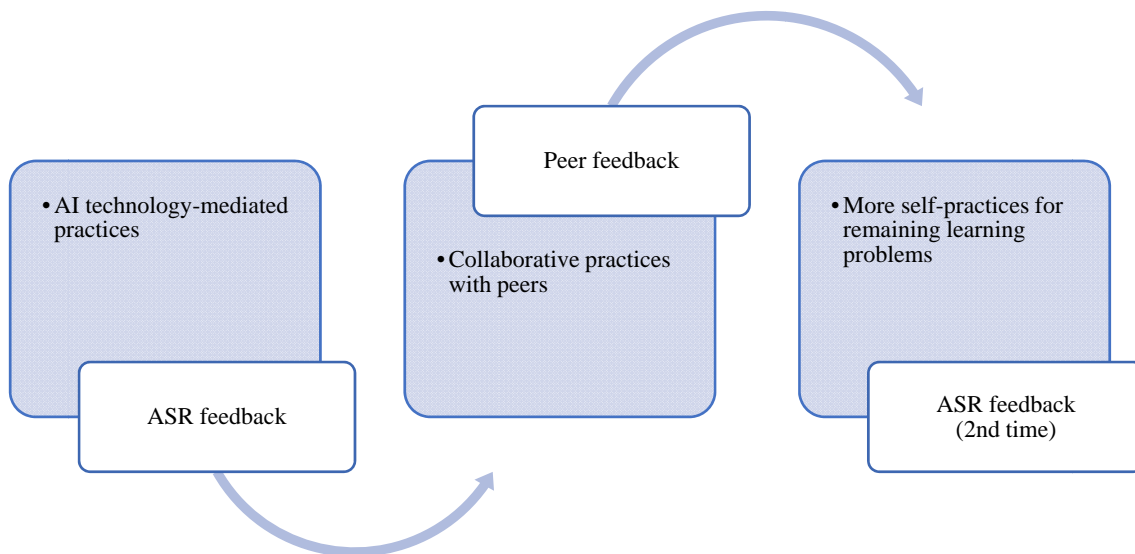


Figure 4. The “sandwich model” of self-regulated learning strategies.

To conclude, clearly determined skill-compatible teaching and learning norms and SRL strategies along with publications of MALL shed light on good mobile learning outcomes in English speaking, especially for regional learners.

### Conclusion

In conclusion, MALL is effective in advancing Chinese college learners' English language informal education. Specifically, IELTS Liulishuo evaluating its seven-dimensional criteria ensures quality intelligent education. However, learners may not adopt it quickly since they prefer the traditional in-person teaching mode and peer feedback. Some difficulties in software and hardware in MALL to meet learners' needs also cause detrimental effects. Therefore, it is recommended that versatility of functions, clear learning goals, achievable independent learning, and intrinsic motivation needs to be considered. More importantly, promoting the status of ASR apps is essential to change negative perceptions towards mobile learning for those who live in regional areas to obtain educational equality.

Undeniably, this literature review has some limitations. One is generalisability of data results. Since the review only adopts one unit of analysis dependent on the IELTS Liulishuo's seven-dimension evaluation framework, it ignores other ASR-enabled apps or oral English tests. Future reviewing should concentrate on the

scope of different spoken-English assessments in MALL, and China's English college tests. Another is the lack of depth in dynamic factors such as *behavior intention* affecting learners' English performances. Future research will further investigate how to promote learners' continued intent on these applications to achieve their long-term English learning goals.

### References

- Ahmadi, H., & Nazari, O. (2014). Grading and sequencing tasks in task-based syllabus: A critical look at criterion selection. *International Journal of Cognitive and Language Sciences*, 8(1), 337-341.
- Brown, H. D. (2004). Grading and student evaluation. In *Language assessment: Principles and classroom practices* (pp. 281-302). New York: Longman.
- Carr, N. T. (2011). What are we testing and why. In *Designing and analyzing language tests* (pp. 5-23). Oxford: Oxford University Press.
- Dai, Y. J., & Wu, Z. W. (2021). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study. *Computer Assisted Language Learning*. Retrieved from <https://doi.org/10.1080/09588221.2021.1952272>
- Deloitte. (2019). *Global development of AI-based education*.
- Fu, S. X., Gu, H. M., & Yang, B. (2020). The affordances of AI-enabled automatic scoring applications on learners' continuous learning intention: An empirical study in China. *British Journal of Educational Technology*, 51(5), 1674-1692.
- Fulcher, G. (2003). Tasks for second language speaking tests. In *Testing second language speaking* (pp. 72-109). London and New York: Routledge.
- IELTS Liulishuo. (2020). v.3.4.1. mobile app. One Plus Market.
- Liu, Y., Yang, T., & Xu, S. Y. (2019). The efficacy of using Liulishuo for spoken English ability—taking the practice of Liulishuo in North China Electric Power University (Baoding) as an example. *International Education Studies*, 12(4), 244-252.
- Sun, Y. Y., & Gao, F. (2020). An investigation of the influence of intrinsic motivation on students' intention to use mobile devices in language learning. *Educational Technology Research and Development*, 68(3), 1181-1198.
- Wang, Y. H., Luan, H., Yuan, J. H., Wang, B. & Lin, H. (2020). LAIX corpus of Chinese learner English: Towards a benchmark for L2 English ASR. In *Interspeech* (pp. 414-418). October 25-29, 2020, Shanghai, China.
- Zou, B., Yan, X. X., & Li, H. (2020). Students' perspectives on using online sources and apps for EFL learning in the mobile-assisted language learning context. In *Language learning and literacy: Breakthroughs in research and practice* (pp. 515-531). Pennsylvania: IGI Global.