

A New Way of Learning Uyghur Language on Mobile Devices

Short Video Interpretation of Intermediate Uyghur Language Listening and Speaking Textbooks

Xingya Wang¹, Amanguli Yiming², Fan Gu³

^{1,2,3} School of Chinese Language, Literature and Art, Xinjiang Agricultural University, Urumqi, China

²Corresponding author.

ABSTRACT

The deep integration of mobile Internet and short video technology is reshaping the ecological pattern of language learning. This article systematically explores the construction logic of the mobile short video deduction mode and the implementation path of Uyghur language learning from three dimensions: mobile learning theory, short video education applications, and multimodal teaching design. The research proposes a four-dimensional design concept of "fragmented input, systematic construction, scenario-based immersion, and interactive participation", and constructs a three-stage teaching loop of "pre-class preheating, in class deepening, and post-class consolidation", providing a replicable theoretical framework and practical paradigm for the digital transformation of language courses. Research has shown that the short video deduction mode effectively reduces cognitive barriers, stimulates learning motivation, and achieves complementary integration between mobile learning and traditional classrooms through multi-sensory collaboration and contextual embedding.

Keywords: Mobile learning, Short videos, Fragmented learning, Multimodal teaching, Blended teaching.

1. INTRODUCTION

1.1 Research Background

Since the 21st century, the exponential development of mobile Internet technology has profoundly changed the way human beings acquire knowledge. The popularization of intelligent terminals has broken through the temporal and spatial boundaries of traditional classrooms, and the mobile learning concept of learning anytime, anywhere is gradually moving from ideal to reality. Statistics released by China Internet Network Information Center show that the number of mobile Internet users in China has exceeded one billion, and the average daily use time of short video applications has continued to rise. The cross-border integration of "short video + education" has become an irreversible trend of the times. In this macro context, the field of language teaching is facing dual pressures of change. On the one hand, the

traditional classroom teaching model is difficult to meet the personalized and differentiated learning needs of learners, and the "three unifications" model of unified progress, unified content and unified method increasingly shows its limitations; On the other hand, the particularity of language learning - emphasizing situational immersion, pronunciation imitation, and communicative practice - places higher demands on teaching media. Simple text textbooks or audio materials are difficult to reproduce real language communication scenarios, and learners often fall into a "mute" knowledge accumulation without context, unable to achieve effective transformation from knowledge to ability. Short videos, as an emerging medium, provide a new technological path for language teaching due to their short duration, fast dissemination, strong interaction, and deep immersion. Compared to traditional multimedia courseware, short videos can more vividly reproduce language communication scenes through multimodal collaboration of visual images, auditory

input, and text subtitles, enabling learners to complete efficient language input and imitation exercises in fragmented time segments. At the same time, the demand for digital teaching of minority languages and foreign languages is becoming increasingly urgent, and how to use modern technological means to improve teaching efficiency has become an important issue in educational research.

1.2 Research Purpose and Significance

The theoretical significance of this study lies in exploring the deep integration path of mobile learning theory and short video media, enriching the application research of multimodal teaching theory in the field of language learning. By systematically analyzing the behavioral characteristics of mobile learning, the knowledge dissemination mechanism of short videos, and the principles of multimodal cognitive processing, a short video learning model framework with theoretical foundation and practical guidance value is constructed. The practical significance lies in providing a replicable and promotable mobile learning paradigm for the teaching reform of language courses through a systematic resource development concept and Uyghur language teaching implementation path design. The research results can be directly applied to the digital transformation of language listening and speaking courses, and provide methodological reference for the development of mobile learning resources in other languages and other disciplines, and help the implementation of education digital strategy.

1.3 Core Research Questions

This study focuses on the following core questions: First, how does short video deduction mode improve language learning effect through multimodal design? Second, how can fragmented learning on the mobile terminal complement traditional teaching effectively and avoid the trap of “fragmented knowledge”? Thirdly, what are the effects of scene narrative and interactive design on learners' cognitive processing and learning motivation?

2. LITERATURE REVIEW AND THEORETICAL BASIS

2.1 Current Research Status of Mobile Learning

Mobile learning refers to the learning activities carried out by learners at any time and any place with the help of mobile terminals. This concept was first proposed by the Irish education technology expert Gideon. After more than 20 years of development, it has evolved from the early simple migration of "moving learning content to mobile phones" to the deep transformation of "learner-centered, context-based, and interaction-centered". Chinese and foreign scholars' research on mobile language learning shows that mobile learning has three advantages: time flexibility. Learners can use scattered periods such as commuting, waiting, and rest for micro-learning, which significantly improves time utilization efficiency. Due to the diversity of scenes, mobile learning can be embedded in real life scenes to achieve “learning by doing” situational learning; the richness of resources and the massive digital resources enable learners to make independent choices based on their personal level and interests to meet differentiated needs. However, mobile learning also faces challenges that cannot be ignored. The first is the problem of distraction. The openness of the mobile environment makes learners vulnerable to the interference of notification pop-ups, social information, etc., resulting in a decrease in learning concentration. The second is the fragmentation of the knowledge system. If there is a lack of systematic knowledge integration mechanism, mobile learning is easy to become a "dragonfly" type of shallow contact, and it is difficult to form a structured knowledge network. The third is that there is a problem of insufficient learning depth. Although short and concise micro-content is easy to absorb, it may weaken learners' ability to think deeply and explore systematically. How to design high-quality mobile learning resources and realize the balance between fragmented learning and systematic knowledge construction is the focus of current research. This study believes that the key to solving this contradiction lies in “systematic fragmentation” — that is, embedding systematic knowledge structure in fragmented presentation forms, and ensuring learners' flexibility at the micro level and integrity at the macro level through mechanisms such as theme modular design, knowledge map association, and learning path guidance.

2.2 Research on the Application of Short Videos in the Field of Education

The knowledge dissemination mechanism of short video can be analyzed from three dimensions: communication, pedagogy and psychology. From the perspective of communication, short video follows the logic of "attention economy", realizes the "hook effect" by maximizing the information density in the first 3 seconds, and quickly captures the users' cognitive resources. From the perspective of pedagogy, short videos disassemble complex knowledge systems into digestible micro-units, reducing the cognitive load of simple learning and meeting the capacity limits of working memory. From the perspective of psychology, the multimodal stimulation of short video can activate multiple perceptual regions of the brain and promote the deep coding and emotional resonance of information. The relationship between fragmented learning and deep learning has been the focus of academic circles for a long time. Critics believe that fragmented learning can easily lead to problems such as weak knowledge mastery, unsystematic learning content, and lack of in-depth thinking training, leaving learners in the dilemma of "knowing a lot and understanding very little". Proponents of argue that fragmented learning is not the opposite of deep learning, but its preconditions and auxiliary means — to establish knowledge contacts through fragmented extensive dabbling, and then to achieve deep understanding through systematic integration and reflection. The two can form a benign interaction of "scattered first and then gathered".

This study holds a dialectical position, and believes that the effect of fragmented learning depends on two key variables: the first is the "connectivity" between fragments, that is, whether there is a clear knowledge context to connect scattered micro-units into an organic whole; the second is the learner's "integration ability", that is, whether there are metacognitive strategies to actively construct fragmentary information into systematic knowledge. Therefore, the design of short video resources should not stop at the content optimization of a single video, but should pay more attention to the logical association and overall architecture between videos. Scene narrative is the core advantage of short video education application. The language knowledge in traditional textbooks is often presented in the form of abstract rules. Learners need to construct their own application scenarios in their minds. This "decontextualization"

process increases the cognitive burden. Short video embeds language knowledge into specific context through real or simulated communication scenarios, so that learners can "see the real environment of language use" and realize contextual understanding of knowledge. Studies have shown that situational embedding can activate learners' prior experience, promote the connection of old and new knowledge, and enhance the transferability of knowledge.

2.3 Multimodal Teaching Theory

The theory of multimodal teaching originates from the cross-study of social semiotics and cognitive science, emphasizing that the essence of human cognition is a multi-channel and multi-sensory collaborative processing process. In language teaching, multimodal presentation is not only a technical means, but also a teaching strategy in line with cognitive rules. Cognitive load theory provides a key perspective for understanding the mechanism of multimodal design. The theory points out that the capacity of human working memory is extremely limited, and the number of information units processed is usually less than 7 ± 2 . When the information is presented improperly - for example, the visual channel is overloaded and the auditory channel is idle - it will generate external cognitive load and crowd out the cognitive resources that are applied to knowledge construction. The value of multimodal design is to achieve "best use and synergy" by rationally allocating the cognitive load of visual, auditory, and textual channels, reduce the pressure of a single channel, and promote the deep coding and long-term maintenance of information. In Uyghur language teaching, the specific mechanism of multimodal presentation is as follows: video images provide situational schemas to help learners establish a scene framework for language use; audio provides standard pronunciation demonstration, so that learners can carry out accurate auditory imitation; subtitles provide text anchors to convert speech streams into recognizable text symbols, reducing the difficulty of auditory decoding. The synergistic effect of the three helps learners to establish a multi-dimensional connection of "pronunciation – form - meaning" and form a solid psychological dictionary. Interactive design is an extension dimension of multimodal teaching. Traditional video viewing is a one-way information receiving behavior, and learners are in a passive position. Interactive design transforms learners from "bystander" to "participant" by embedding pause thinking, follow-up imitation, instant test,

comment communication and other links, and activates their subjectivity and sense of engagement. Studies have shown that interactivity can not only enhance learners' participation, but also promote the internalization and transformation of knowledge through "output-driven".

2.4 Research Review and Positioning of This Study

Based on existing research, the combination of mobile learning and short video education has a solid theoretical foundation, but there is relatively insufficient systematic design research for language teaching. Previous studies have mostly focused on case applications or effect descriptions of short videos, lacking a complete framework from theoretical construction to pattern design and implementation path. This study attempts to achieve breakthroughs in the following areas: Firstly, the four-dimensional design concept of "fragmented input, systematic construction, scenario immersion, and interactive participation" is proposed to organically unify the fragmented presentation form with the systematic knowledge structure; Secondly, establish a three-stage teaching loop of "pre class warm-up, in class deepening, and post class consolidation", and clarify the complementary mechanism between mobile learning and traditional classrooms; Thirdly, starting from the theory of cognitive load, analyze the specific role path of multimodal design in language learning, and provide theoretical basis for resource development.

3. THEORETICAL CONSTRUCTION OF SHORT VIDEO LEARNING MODE

3.1 Four-dimensional Design Concept

Based on literature review and theoretical analysis, this study proposes a four-dimensional design concept for short video learning mode: Fragmented input: The audio and video duration of each micro unit in the intermediate Uyghur language listening and speaking course is controlled at 3-5 minutes to adapt to the physiological limitations of learners' attention duration, making it easier to complete learning during breaks, commuting, waiting, and other scattered periods. Fragmentation is not a random cut of content, but a "meaningful split" based on the internal logic of the knowledge module to

ensure that each micro-unit is independent and complete, and is associated with other units.

Systematic construction: On the basis of fragmented presentation forms, a knowledge network of "theme-led, unit-progressive" is constructed. The 32 major themes cover the core areas of daily life, campus learning, and social interaction. Each theme has 3-4 micro-units, forming a gradient structure from basic to advanced, from simple to complex. Learners can not only learn systematically according to the order of topics, but also jump to choose according to their personal needs, so as to achieve "flexibility in structure".

Scene-based immersion: Through the restoration of real communication scenes, learners are placed in an "immersive" language environment. Scene design follows the principles of "authenticity" and "typicality", which not only reflects the real context of language use, but also refines the most representative types of communication. Scene narration not only provides language input, but also conveys tacit knowledge such as cultural background, communicative strategies and pragmatic rules, so as to promote the development of learners' pragmatic competence.

Interactive participation: It is to embed interactive links such as pause follow-up reading and imitation exercises in Uyghur learning videos, and change "passive viewing" to "active participation". The interactive design follows the principle of "immediate feedback", so that learners can timely test the learning effect, correct the wrong cognition, and form a virtuous circle of "learning - feedback - adjustment". At the same time, the supporting learning community supports the communication and sharing among learners and builds a collaborative learning ecology.

3.2 Resource System Architecture

The resource system adopts the architecture design of "dual-track parallel and module combination". The video track focuses on scene demonstration and visual assistance, and restores the communicative scene through multi-dimensional information such as pictures, characters, movements and expressions; the audio track focuses on pronunciation training and listening enhancement, stripping visual interference, so that learners can focus on speech recognition and imitation. Uyghur learners can choose flexibly according to their learning objectives and scene needs - visual learners prefer video tracks, auditory

learners prefer audio tracks, and deep learners can be used in combination to achieve double reinforcement. The supporting learning support system provides three core functions: progress tracking, recording learners' learning trajectory, completion rate, repeated viewing times and other data, generating personal learning portraits; error review, automatic marking of learners' errors in the test, push targeted review resources; community discussion, support learners to communicate around specific topics, share learning experiences and questions, and form a learning atmosphere of peer assistance.

3.3 Analysis of Cognitive Mechanisms

From the perspective of cognitive science, the effectiveness of short video learning mode stems from the synergy of three mechanisms: multi-sensory synergy promotes memory coding. Visual pictures activate the visual cortex of the occipital lobe, auditory input activates the auditory cortex of the temporal lobe, and text reading activates the language area of the parietal lobe. The synergy of the three activates a wide area of the brain and promotes multi-channel coding of information. According to the dual coding theory, the representation of the verbal system and the non-verbal system at the same time can significantly improve the retention rate and retrieval rate of memory. Situational embedding lowers the cognitive threshold. The abstract language rules in traditional textbooks are transformed into perceptible and understandable communicative behaviors through the concrete presentation of real scenes, which reduces learners' psychological resistance and cognitive load. Situation not only provides understanding clues, but also stimulates learners' emotional resonance, so that the learning process shifts from "cold cognition" to "hot experience". Instant accessibility improves learning frequency. The convenient access on the mobile terminal eliminates the time and space obstacles of learning, so that learning behavior can be embedded in the high-frequency scenes of daily life. According to the "distributed practice effect", learning is dispersed into multiple time periods, which is better than centralized learning with the same total duration. Fragmentation design is the use of this cognitive law, through the accumulation of "a small number of times" to achieve quantitative change to qualitative change.

4. DESIGN OF TEACHING IMPLEMENTATION PATH

4.1 The Three-stage Teaching Closed-loop

The short video learning model is not an alternative to the traditional classroom, but a three-stage complementary path of "pre-class - in-class - post-class" to construct a blended teaching ecology integrating online and offline. In the pre-class warm-up stage, learners watch short videos through mobile phones, initially contact new topics and core expressions, and complete "input warm-up". The goal of this stage is to establish a prior knowledge framework, activate relevant background experience, and prepare for Uyghur classroom learning. Teachers can understand learners' preview through platform data, identify common difficulties and individual differences, and provide a basis for classroom design. In the deepening stage of the lesson, based on the learner's preview data, the teacher focuses on the difficulties to carry out intensive lectures and interactive exercises to achieve "precise teaching". The classroom time has shifted from "knowledge imparting" to "ability training", focusing on high-level activities such as situational simulation, role-playing, and group discussion to promote the internalization and transformation of knowledge. The role of teachers has changed from "knowledge authority" to "learning designer" and "ability coach". In the post-class consolidation stage, learners repeatedly watch videos, complete follow-up exercises, submit homework through the platform, participate in discussions, and complete "output enhancement". This stage emphasizes the automation and proficiency of knowledge, and forms language intuition through repeated practice. At the same time, learners can check the missing problems in the classroom to achieve personalized review.

4.2 Transformation Teacher's Role

In the short video learning mode, the role of teachers has undergone a triple transformation: from "content lecturer" to "learning designer", responsible for the planning, development, screening and integration of short video resources; from "knowledge authority" to "learning facilitator", they organize interactive activities in the classroom, provide instant feedback, and stimulate deep thinking; from "experience dependent" to "data driven", through the data analysis of the learning platform, they understand the learning situation, so as to achieve accurate

teaching. Teachers' professional development needs corresponding ability improvement: the first is the short video Uyghur teaching design ability, including script writing, scene design, interactive embedding, etc. the second is the data analysis ability, including interpreting learning behaviors, evaluating learning outcomes, and adjusting teaching strategies; the third is the organizational ability of blended learning, including the connection between online and offline activities, coordination of synchronous and asynchronous learning, etc.

4.3 Learner Ability Development

The short video learning mode has put forward new ability requirements for learners. Firstly, there is the ability for self-directed learning, where learners need to possess metacognitive skills such as goal setting, time management, progress monitoring, and strategy adjustment to avoid losing direction in abundant resources. Secondly, there is the ability to integrate information. Learners need to actively construct fragmented micro units into systematic knowledge structures, avoiding the situation of "seeing only the trees but not the forest". Thirdly, it is the digital literacy. Learners need to be familiar with the functional operation of mobile learning platform, and have the ability of information screening and resource screening. Teachers bear an important responsibility in cultivating the ability of Uyghur learners: to reduce the cognitive load of autonomous learning through clear learning path guidance, promote knowledge integration through regular learning reflection activities, and improve the efficiency of learners' platform use through digital literacy training.

5. ADVANTAGES, LIMITATIONS, AND OPTIMIZATION STRATEGIES OF THE MODEL

5.1 Comparative Advantage Analysis

Compared with the traditional teaching mode, the short video deduction mode shows three comparative advantages:

Flexibility advantage: Breaking the limitation of fixed class hours and fixed places, Uyghur learners can arrange their learning progress independently according to their personal rhythm, so as to realize "my learning is my decision". This flexibility not only improves the efficiency of time utilization, but

also respects the individual differences and individual needs of learners.

Controllable advantage: the playback feature of video enables learners to repeatedly observe and imitate difficulties until they master them. This "repeatability" makes up for the lack of "transient" classroom teaching and provides a fair learning opportunity for learners with different learning speeds.

Consistency advantage: Standardized pronunciation demonstration and standardized expression ensure that the input quality received by all learners is consistent, and reduce the fluctuation of learning effect caused by teachers' individual differences (such as accent, speech speed, expression habits). This is particularly important for language beginners and can help them establish a correct phonetic basis.

5.2 Reflection on Existing Limitations

Although the short video mode has significant advantages, the study also reveals several limitations:

The limitations of scene expressiveness: Limited by factors such as shooting conditions, production costs, and actor level, the realism and richness of some short video scenes are insufficient, and it is difficult to fully restore complex communication situations. How to improve the performance of the scene within the limited budget is a realistic challenge for resource development.

The limitations of learning depth: Although fragmented design reduces the cognitive load of single learning, it may also weaken the opportunity for learners to think systematically and explore deeply. How to find a balance between fragmentation and depth requires more detailed teaching design.

The limitations of social interaction: Short video learning is essentially an individualized behavior. If there is no effective community operation and collaboration mechanism, it will easily lead to the increase of learners' loneliness and loss rate. How to build a virtual Uyghur language learning community and promote communication and mutual assistance among learners is an important direction for optimizing the model.

5.3 Optimization Strategy Suggestions

In response to the above limitations, this study proposes the following optimization strategies: The first is to enhance the authenticity of the scene: It is necessary to introduce virtual reality (VR) or augmented reality (AR) technology to construct immersive 3D communication scenes, invite native speakers to participate in filming to enhance the authenticity of language use and adopt documentary style filming techniques to enhance the realism and immersion of the scene. The second is to deepen knowledge integration: At the end of each theme module, there is a must to set up a "knowledge sorting" section to help learners establish their knowledge structure through visual tools such as mind maps and concept maps, and there is also a must to design comprehensive tasks across units to promote the transfer and application of knowledge. The third is to strengthen community interaction: It is also necessary to establish a Uyghur language learning check-in mechanism to enhance learning persistence through peer pressure and positive motivation, and set up a themed discussion area to encourage learners to share their learning experiences and questions. Meanwhile, it is a must to regularly organize online learning activities, such as dubbing competitions, situational drama performances, etc., to enhance the fun and sense of belonging in learning.

6. CONCLUSION

The mobile short video interpretation of intermediate Uyghur language listening and speaking textbooks is an innovative exploration of the integration of mobile learning theory and multimodal teaching practice. In the future, with the integration of virtual reality and artificial intelligence technology, the short video learning mode will achieve greater breakthroughs in scene authenticity and personalized adaptation, continuously helping to implement digital education strategies and cultivate compound language talents.

ACKNOWLEDGMENTS

Fund project: This paper is a phased achievement of the research on Uyghur language phonetics teaching under the background of excellent traditional Chinese culture (project number: 2024PTJG146). This article is a phased achievement of the Autonomous Region level College Students' Innovation and Entrepreneurship

Training Program project "A New Way of Learning Uyghur Language on Mobile Devices - Short Video Interpretation of Intermediate Uyghur Listening and Speaking Textbooks" (Project Number: S202510758069).

REFERENCES

- [1] Chen Huanliang, Yu Qi, Tu Wenzhi, Research on the Current Situation of Fragmented Learning Among Medical Graduate Students [J]. Chinese Journal of Medical Education, 2020, 40(10): 804-807.
- [2] Wang Mi, On Designing the Content of Micro-Video Courses in a Fragment Learning Age [D]. Shanghai: East China Normal University, 2013.
- [3] Huang Jianfeng, Fragmented Learning: New Learning Style Based on "Internet Plus" [J]. Education Exploration, 2016(12): 115-119.
- [4] Wei Xuefeng, Yang Xianmin, Zhang Yuhai, Application Scenarios and Efficient Management of Fragmented Learning Resources in Mobile Internet Era [J]. China Educational Technology, 2017(364): 117-122.
- [5] Li Tingsong, Zhang Hongsu, Lv Linhai, Advantages and Disadvantages of New Learning Style in New Age — Current Situation, Problems and Countermeasures of Fragmented Learning of College Students in the Internet Age [J]. Research in Teaching, 2020, 43(2): 1-5.
- [6] Wang Zhuli, Fragmentation of Learning and Its Countermeasures in Mobile Internet Era: From "Fixed Deposit by Installments" Learning Strategies to "Internet +" Classroom [J]. Journal of Distance Education, 2016, 34(4): 9-16.
- [7] Peng Jingwen, Xu Xiangyun, Fixing the Issue of "Fragmented Learning": Reflection on MOOC Learning Support [J]. Jiangsu Higher Education, 2017(5): 42-46.
- [8] Mayer R E. The Cambridge Handbook of Multimedia Learning[M]. Cambridge: Cambridge University Press, 2009.