Innovation Humanities and Social Sciences Research (IHSSR)
Volume 22, Issue 4(2026). ISSN：2949-1282

A New Way of Learning Uyghur Language on Mobile Devices
Short Video Interpretation of Intermediate Uyghur Language Listening and Speaking Textbooks

Xingya Wang[1], Amanguli Yiming[2], Fan Gu[3]

[1,2,3] *School of Chinese Language, Literature and Art, Xinjiang Agricultural University, Urumqi, China*
[2]*Corresponding 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 references for the development of mobile learning resources for other languages and disciplines, helping to implement the digital education strategy.

1.3 Core Research Questions

This study focuses on the following core questions: Firstly, how can the short video deduction mode improve language learning effectiveness through multimodal design? Secondly, how can fragmented learning on mobile devices effectively complement traditional teaching and avoid the trap of "fragmented knowledge"? Thirdly, what impact does scenario based narrative and interactive design have 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 conducted by learners at any time and any place using mobile terminals. This concept was first proposed by Irish educational technology expert Keegan, and after more than 20 years of development, it has evolved from a simple transfer of learning content to mobile phones in the early days to a deep transformation centered on learners, based on context, and centered on interaction. Chinese and foreign scholars' research on mobile language learning shows that mobile learning has three advantages: time flexibility, learners can use scattered time periods such as commuting, waiting, and rest for micro learning, significantly improving time utilization efficiency; Diversity of scenarios, mobile learning can be embedded in real life scenarios to achieve contextualized learning through “learning by doing”; Abundant resources and massive digital resources enable learners to make independent choices based on their personal level and interests, meeting differentiated needs. However, mobile learning also faces significant challenges that cannot be ignored. Firstly, there is the issue of attention dispersion. The openness of mobile environments makes learners susceptible to interference from notification pop ups, social information, and other factors, leading to a decrease in learning focus. Secondly, there is the problem of fragmented knowledge systems. Without a systematic mechanism for knowledge integration, mobile learning is prone to become shallow and superficial contacts, making it difficult to form a structured knowledge network. Thirdly, there is a problem of insufficient learning depth. Although short and concise micro content is easy to absorb, it may weaken learners' ability to engage in deep thinking and systematic exploration. How to design high-quality mobile learning resources and achieve a balance between fragmented learning and systematic knowledge construction is currently a key research direction. This study believes that the key to resolving this contradiction lies in "systematic fragmentation" — embedding systematic knowledge structures 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 graph 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 videos can be analyzed from three dimensions: communication studies, education, and psychology. From the perspective of communication studies, short videos follow the logic of "attention economy" and achieve the "hook effect" by maximizing the information density in the first 3 seconds, quickly capturing users' cognitive resources. From an educational perspective, short videos break down complex knowledge systems into digestible micro units, reducing the cognitive load of a single learning session and meeting the capacity limitations of working memory. From a psychological perspective, multimodal stimulation of short videos can activate multiple perceptual regions of the brain, promoting deep encoding of information and emotional resonance. The relationship between fragmented learning and deep learning has long been a focus of attention in academia. Critics believe that fragmented learning can easily lead to problems such as weak knowledge acquisition, unsystematic learning content, and insufficient thinking training, causing learners to fall into the dilemma of "knowing a lot but understanding very little". Supporters believe that fragmented learning is not the opposite of deep learning, but rather its prerequisite and auxiliary means — establishing knowledge touchpoints through extensive exploration of fragmented learning, and then achieving deep understanding through systematic integration and reflection. The two can form a virtuous interaction of "dispersion before aggregation".

This study takes a dialectical stance and believes that the effectiveness of fragmented learning depends on two key variables: first, the "connectivity" between fragments, that is, whether there is a clear knowledge network that connects dispersed micro units into an organic whole; The second is the learner's "integration ability", which refers to whether they have metacognitive strategies to actively construct fragmented information into systematic knowledge. Therefore, the design of short video resources should not only focus on optimizing the content of individual videos, but also pay attention to the logical connections and overall architecture between videos. Scene-based narrative is the core advantage of short video education applications. The language knowledge in traditional textbooks is often presented in the form of abstract rules, and learners need to construct their own application scenarios in their minds. This "de contextualization" process increases cognitive burden. Short videos embed language knowledge into specific contextual contexts through real or simulated communication scenarios, enabling learners to "see the real environment of language use" and achieve contextualized understanding of knowledge. Research has shown that situational embedding can activate learners' prior experience, promote the connection between new and old knowledge, and enhance the transferability of knowledge.

2.3 Multimodal Teaching Theory

The multimodal teaching theory originates from the interdisciplinary research of social semiotics and cognitive science, emphasizing that the essence of human cognition is a collaborative processing process involving multiple channels and senses. In language teaching, multimodal presentation is not only a technical means, but also a teaching strategy that conforms to cognitive laws. The cognitive load theory provides a key perspective for understanding the mechanisms of multimodal design. This theory suggests that the capacity of human working memory is extremely limited, and the number of information units processed simultaneously typically does not exceed 7 ± 2. When the presentation of information is inappropriate, such as visual channel overload and auditory channel idleness, external cognitive load will be generated, occupying cognitive resources that should be used for knowledge construction. The value of multimodal design lies in the rational allocation of cognitive load across visual, auditory, textual, and other channels, achieving "each doing its best and synergistic", reducing the pressure on a single channel, and promoting deep encoding and long-term preservation of information. In Uyghur language teaching, the specific mechanism of multimodal presentation is as follows: video images provide situational schemata, helping learners establish a contextual framework for language use; Audio provides standard pronunciation demonstrations, enabling learners to perform 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 establish a multidimensional connection of "sound form meaning" and form a stable psychological dictionary. Interactive design is an extended dimension of multimodal teaching. Traditional video watching is a one-way

information receiving behavior, and learners are in a passive position. Interactive design transforms learners from "bystanders" to "participants" by embedding pause thinking, imitation reading, real-time quizzes, and comment communication, activating their subjectivity and sense of engagement. Research has shown that interactivity not only enhances learners' engagement, but also promotes the internalization and transformation of knowledge through output driven learning.

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 the arbitrary cutting of content, but rather the "meaningful splitting" based on the inherent logic of knowledge modules, ensuring that each micro unit is both independent and complete, while also forming connections with other units. Systematic construction: It is to construct a knowledge network with "theme leadership and unit progression" on top of fragmented presentation forms. The 32 major themes cover core areas such as daily life, campus learning, and social interaction, with 3-4 micro units set up under each theme, forming a gradient structure from basic to advanced, and from simple to complex. Learners can learn systematically according to the order of topics, or skip learning according to personal needs, achieving "flexibility in structure". Scene-based immersion: By reproducing real-life communication scenarios, learners are immersed in an immersive language environment. Scene design follows the principles of "authenticity" and "typicality", reflecting both the real context of language use and extracting the most representative types of communication. Situational narrative not only provides language input, but also conveys implicit knowledge such as cultural background, communication strategies, and pragmatic rules, promoting the development of learners' pragmatic abilities. Interactive participation: Embedding interactive elements such as pause and imitation exercises in Uyghur language learning videos, transforming "passive viewing" into "active participation". Interactive design follows the principle of "instant feedback", enabling learners to timely verify learning effectiveness, correct erroneous cognition, and form a virtuous cycle of "learning-feedback-adjustment". At the same time, the supporting learning community supports communication and sharing among learners, building a collaborative learning ecosystem.

3.2 Resource System Architecture

The resource system adopts a dual track parallel and module combination architecture design. The video track focuses on scene demonstration and visual assistance, restoring communication scenes through multi-dimensional information such as images, characters, actions, and expressions; The audio track focuses on pronunciation training and listening enhancement, removing visual interference and allowing learners to focus on speech recognition and imitation. Uyghur language learners can flexibly choose based on their learning objectives and situational needs - visual learners

prefer video tracks, auditory learners prefer audio tracks, and deep learners can use them in combination to achieve dual reinforcement. The supporting learning support system provides three core functions: the first is progress tracking, recording learners' learning trajectories, completion rates, repeated viewing times, and generating personal learning profiles. The second is error review, automatically marking learners' mistakes in the test and pushing targeted review resources. The third is community discussions support learners to engage in communication around specific topics, share learning experiences and questions, and create a learning atmosphere of peer assistance.

3.3 Analysis of Cognitive Mechanisms

From the perspective of cognitive science, the effectiveness of short video learning models stems from the synergistic effect of three mechanisms: Multi-sensory collaboration promotes memory encoding. Visual images 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 range of brain regions, promoting multi-channel encoding of information. According to the dual coding theory, obtaining representations of both verbal and nonverbal systems can significantly improve memory retention and retrieval rates. Context embedding lowers cognitive barriers. The abstract language rules in traditional textbooks are presented in a concrete way through real-life scenarios, transforming them into perceivable and understandable communicative behaviors, reducing learners' psychological resistance and cognitive load. Scenarios not only provide clues for understanding, but also stimulate learners' emotional resonance, shifting the learning process from "cold cognition" to "hot experience". Instant accessibility enhances learning frequency. The convenient access on mobile devices eliminates time and space barriers for learning, allowing learning behavior to be embedded in high-frequency scenarios of daily life. According to the "distributed practice effect", dispersing learning across multiple time periods is more effective than centralized learning of the same total duration. Fragmented design utilizes this cognitive law to achieve the transformation from quantitative change to qualitative change through the accumulation of "a small amount of multiple times".

4. DESIGN OF TEACHING IMPLEMENTATION PATH

4.1 The Three-stage Teaching Closed-loop

Short video learning model is not a replacement for traditional classrooms, but rather forms a three-stage complementary path of "pre class, in class, and after class", constructing a blended learning ecosystem that integrates online and offline teaching. During the pre-class warm-up stage, learners watch short videos on their mobile devices to gain initial exposure to new topics and core expressions, completing the "input warm-up". The goal of this stage is to establish a prior knowledge framework, activate relevant background experience, and prepare cognitively for Uyghur language classroom learning. Teachers can use platform data to understand learners' preview situations, identify common difficulties and individual differences, and provide a basis for classroom design. In the deepening stage of class, teachers focus on difficult points and conduct precise lectures and interactive exercises based on learners' preview data, achieving "precision teaching". The classroom time has shifted from "knowledge imparting" to "ability training", with a focus on high-level activities such as situational simulation, role-playing, and group discussions to promote the internalization and transformation of knowledge. The role of teachers has shifted from "knowledge authority" to "learning designer" and "ability coach". In the post class consolidation stage, learners repeatedly watch videos, complete reading exercises, submit assignments through the platform, participate in discussions, and complete "output reinforcement". This stage emphasizes the automation and proficiency of knowledge, forming language intuition through repeated practice. At the same time, learners can identify and address any remaining issues in the classroom, achieving personalized review.

4.2 Transformation Teacher's Role

In the short video learning mode, the role of teachers undergoes a triple transformation: from "content lecturer" to "learning designer", responsible for planning, developing, screening, and integrating short video resources; The teachers transform from a "knowledge authority" to a "learning facilitator", organizing interactive activities in the classroom, providing immediate feedback, and stimulating deep thinking; The teachers transform from an "experience dependent"

to a "data-driven" approach, understanding learning situations through data analysis on learning platforms, and achieving precise teaching. The professional development of teachers requires corresponding ability enhancement: the first is the short video Uyghur language 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, digital literacy requires learners to be familiar with the functional operations of mobile learning platforms and possess the ability to discern information and filter resources. Teachers have an important responsibility in cultivating the abilities of Uyghur language learners: guiding them through clear learning paths to reduce the cognitive load of self-directed learning; Promoting knowledge integration through regular learning and reflection activities; Enhancing learners' platform utilization efficiency through digital literacy training.

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

5.1 Comparative Advantage Analysis

Compared to traditional teaching methods, the short video deduction mode exhibits three comparative advantages: flexibility advantage: Breaking the limitations of fixed class hours and fixed locations, Uyghur language learners can independently arrange their learning progress according to their personal pace, achieving the goal of “I am the master of my own learning”. This flexibility not only improves the efficiency of time utilization, but also respects the individual differences and personalized needs of learners. Controllability advantage: The replayable nature of videos allows learners to repeatedly observe and imitate difficult points until they master them. This "repeatability" compensates for the lack of "one stop shop" in classroom teaching and provides fair learning opportunities for learners with different learning speeds. Consistency advantage: Standardized pronunciation demonstration and standardized expression ensure that all learners receive consistent input quality, reducing learning effect fluctuations caused by individual differences among teachers (such as accent, speed, and expression habits). This is particularly important for language beginners as it can help them establish a correct phonetic foundation.

5.2 Reflection on Existing Limitations

Although the short video mode has significant advantages, research has also revealed several limitations: The limitations of scene expression: Due to limitations such as filming conditions, production costs, and actor skills, the realism and richness of some short video scenes are insufficient, making it difficult to fully reproduce complex communicative situations. How to enhance the expressiveness of scenarios within a limited budget is a practical challenge faced by resource development. The limitations of learning depth: Although fragmented design reduces the cognitive load of single learning, it may also weaken learners' opportunities for systematic thinking and deep exploration. Finding a balance between fragmentation and depth requires more refined instructional design. The limitations of social interaction: Short video learning is essentially an individualized behavior, and without effective community operation and collaboration mechanisms, it can easily lead to learners' feelings of loneliness and increased churn rates. 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.