A Study of Emotion Regulation and Intervention in Second Language Learning at Vocational Undergraduate Level in the Context of Multimodel Large Language Model

Huanhuan Feng[1,a*], Lingli Wang[1,b], Yuting Wu[2,c], Mengmeng Wu[1,d]

[1]School of Foreign Languages, Jingdezhen Vocational University of Art, Jingdezhen, Jiangxi, China
[2]School of Finance, Economics and Law, Jingdezhen Vocational University of Art, Jingdezhen, Jiangxi, China
[a]164352674@qq.com, [b]1476929884@qq.com, [c]515121892@qq.com, [d]1119586576@qq.com
*Corresponding author

Keywords: Emotion Regulation; Multimodal Large Language Models; Vocational Undergraduate Education; Second Language Learning; Affective Pedagogy; AI-Assisted Instruction

Abstract: The emergence of multimodal large language models (MMLLMs)—such as GPT-4, Gemini, and Sora—has introduced a paradigm shift in second language (L2) education, enabling not only cognitive enhancement but also affective support through real-time, multimodal interaction. However, their potential for scaffolding learner emotion regulation remains underexamined, particularly within vocational undergraduate contexts where students often face heightened anxiety, low academic self-concept, and limited emotional self-regulation capacity. This study investigates the emotional characteristics of vocational L2 learners and examines how MMLLMs can be pedagogically leveraged to foster emotionally responsive, context-aware, and professionally relevant instruction. Drawing on a mixed-methods approach involving classroom observation and semi-structured interviews, the findings suggest that MMLLMs, when integrated with targeted strategies—such as emotion-aware adaptation, multimodal content delivery, personalized feedback, and vocational task simulation—significantly enhance both emotional resilience and communicative competence. The study provides theoretical insights and practical implications for developing affect-sensitive, AI-assisted L2 pedagogy tailored to the needs of vocational learners.

1. Introduction

Emotions play a central yet often underestimated role in the process of second language acquisition (SLA). For vocational undergraduate students—who frequently grapple with pragmatic learning motives, limited language exposure, and reduced academic self-concept—emotional factors such as anxiety, boredom, and disengagement present formidable barriers to successful L2

attainment.

In parallel, the recent surge in artificial intelligence (AI), particularly in the form of multimodal large language models (MMLLMs), has introduced novel opportunities for transforming not only cognitive but also affective dimensions of language learning. By enabling dynamic human-computer interaction through text, speech, images, and video, MMLLMs offer an emotionally responsive learning environment, capable of tailoring feedback and engagement strategies to individual learner profiles.

1.1 Learner Emotion Regulation

Emotion regulation refers to the process by which individuals monitor, evaluate, and modify their emotional reactions to achieve personal goals and maintain psychological well-being[1]. In the context of second language (L2) acquisition, learners' ability to regulate emotions plays a critical role in shaping their motivation, engagement, and overall language performance. Negative emotions such as anxiety, frustration, or embarrassment can significantly impair concentration, inhibit language production, and decrease willingness to communicate. Conversely, effective emotion regulation strategies can enhance cognitive functioning, foster resilience, and sustain learning persistence.

According to Gross's process model, emotion regulation involves antecedent-focused and response-focused strategies, including situation selection, cognitive reappraisal, attention deployment, and suppression [2]. In L2 learning environments, cognitive reappraisal—such as viewing mistakes as learning opportunities—has been associated with improved academic resilience and performance. Furthermore, self-regulated learners often employ strategies such as self-talk, goal setting, and emotional awareness to manage language anxiety and maintain a growth-oriented mindset[3].

Vocational undergraduate students, in particular, may face additional emotional challenges due to low self-efficacy in language learning, limited exposure to authentic language environments, or pressure from professional training. Therefore, strengthening emotion regulation capacities among L2 learners is essential for building both linguistic competence and psychological readiness for real-world communication tasks.

With the emergence of artificial intelligence and multimodal technologies, it is now possible to recognize, support, and even scaffold emotion regulation through intelligent systems. As such, learner emotion regulation is not only an internal psychological mechanism but also a vital focus of emotionally responsive and adaptive language instruction in the digital age.

1.2 Multimodal Large Language Modeling for Teaching and Learning

Multimodal large language models (MMLLMs) represent a significant advancement in AI-driven education by integrating multiple input and output modalities—such as text, audio, image, and video—into a unified language understanding and generation framework. Unlike earlier unimodal models that relied solely on textual input, MMLLMs possess the capacity to interpret and generate rich, contextualized content across modalities, thereby facilitating more dynamic, immersive, and personalized learning experiences [4].

In the context of second language learning, MMLLMs offer transformative affordances that extend beyond vocabulary acquisition and grammar instruction. These models can deliver real-time conversational feedback, simulate interactive speaking scenarios, generate culturally nuanced multimodal learning materials, and provide tailored pronunciation guidance through speech synthesis and recognition. Their ability to analyze not only linguistic but also paralinguistic and visual cues makes them particularly effective in enhancing learner engagement and understanding

[5].

From an instructional standpoint, MMLLMs support adaptive pedagogy by enabling emotion-aware feedback, task personalization, and multimodal scaffolding[6]. For example, when learners demonstrate confusion or hesitation in speaking tasks, the system can detect vocal tone changes or facial expressions and respond with simplified input, encouraging prompts, or visual aids. This responsive feedback loop creates a learner-centered environment that is sensitive to emotional states and cognitive needs.

Moreover, MMLLMs enable instructors to co-design authentic, industry-relevant scenarios tailored to vocational learners. For instance, students in business English or ceramic trade programs can engage in AI-mediated simulations of cross-cultural negotiations, product presentations, or trade fair dialogues—complete with multimedia context and real-time language support [7]. These simulations not only foster communicative competence but also cultivate professional confidence.

As educational technologies continue to evolve, the integration of MMLLMs into language instruction signals a shift toward emotionally intelligent, context-aware, and multimodally enriched learning ecosystems. For vocational undergraduate learners, whose needs are often overlooked in mainstream SLA research, MMLLMs present a unique opportunity to bridge the gap between academic English instruction and workplace language application.

2. Emotional characteristics of L2 learners in vocational undergraduate programs

Building on the foundational understanding of learner emotion regulation and the transformative potential of multimodal large language models (MMLLMs) in second language education, it is essential to closely examine the specific emotional characteristics exhibited by vocational undergraduate L2 learners. These learners face unique affective challenges that shape their engagement, motivation, and overall language acquisition trajectories. By identifying and analyzing these emotional dynamics—ranging from anxiety and self-concept issues to motivational misalignments and dependence on external support—educators and researchers can better tailor MMLLM-enhanced interventions to effectively address the complex interplay between cognition and emotion in vocational L2 contexts.

2.1 Heightened Emotional Barriers to L2 Engagement

Vocational undergraduate learners frequently experience heightened negative affective states during their second language acquisition process, with language anxiety being the most prominent. This anxiety is particularly acute in oral communication activities such as speaking and listening, where the fear of negative evaluation by peers or instructors often leads to avoidance behaviors, silence, or overreliance on memorization rather than authentic communicative practice. Additionally, the inherent complexity and rapid pace of classroom tasks can exacerbate learners' emotional vulnerability, causing frustration, cognitive overload, and even disengagement from the learning process. The absence of effective coping or emotion regulation mechanisms often results in a cyclical pattern where negative emotions hinder language practice, which in turn reinforces anxiety and avoidance. This emotional barrier constitutes a significant obstacle to consistent participation and impedes the gradual development of communicative competence in vocational contexts, where practical application and interaction skills are essential.

2.2 Low Academic Self-Concept and Underdeveloped Emotion Regulation

Many vocational undergraduates possess a weakened academic self-concept regarding their L2 abilities, shaped by a history of perceived failures, limited success in standardized assessments, and

internalized negative beliefs about their language learning potential. This diminished self-confidence directly affects learners' willingness to engage in communicative risk-taking, which is crucial for language acquisition. Compounding this issue is the widespread lack of internalized emotion regulation strategies among vocational learners. Unlike their counterparts in traditional academic pathways, these students often exhibit low emotional awareness and find it challenging to identify and manage negative affective responses such as stress, frustration, or embarrassment. They may struggle to reframe setbacks constructively or to sustain intrinsic motivation over time. This deficit highlights a pressing need for explicit and structured emotional regulation support embedded within L2 instruction, enabling learners to develop greater emotional resilience, enhance self-efficacy, and cultivate a growth mindset necessary for overcoming linguistic challenges.

2.3 Reliance on External Support and Misalignment of Motivation

Vocational students generally demonstrate strong instrumental motivation, driven by pragmatic goals such as securing employment, meeting internship requirements, or fulfilling program credits. However, this externally focused motivation does not always correlate with emotional engagement or intrinsic interest in the language learning process. When course content appears disconnected from learners' immediate career objectives or personal interests, they often experience emotional disengagement, boredom, or even resentment towards the learning tasks. Moreover, these students tend to rely heavily on external sources of emotional reinforcement—including teacher praise, peer encouragement, and visible markers of achievement—to maintain motivation and sustain effort. In the absence of consistent, timely, and affectively attuned feedback, their persistence in language learning may wane, leading to decreased participation and plateaued progress. This dynamic underscores the importance of designing affectively responsive learning environments that align curricular content with vocational relevance while providing ongoing emotional support to nurture learner engagement.

3. Personalized Teaching Strategies Based on Multimodal Large Language Models

To effectively address the emotional and cognitive needs of vocational undergraduate L2 learners, personalized teaching strategies must be adopted—strategies that not only recognize learner diversity but also dynamically respond to their affective and instructional profiles. Multimodal large language models (MLLMs) offer powerful capabilities for realizing such personalization. Through the integration of real-time emotional analysis, multimodal content generation, and adaptive learning trajectories, MLLMs enable instructors to create individualized, emotionally supportive, and pedagogically aligned learning environments. The following subsections detail four key strategies that illustrate how MLLMs can be operationalized to enhance language instruction at the vocational undergraduate level. These include: emotion-aware instructional adaptation, multimodal content aligned with learning styles, adaptive feedback coupled with task personalization, and the generation of contextualized output tasks grounded in real-world professional domains[8].

3.1 Emotion-Aware Instructional Adaptation

Multimodal large language models (MLLMs) can detect learners' emotional states—such as anxiety, fatigue, or frustration—through analysis of multimodal cues including speech tone, facial expressions, and linguistic features. By capturing these subtle affective fluctuations, instruction can be dynamically adapted to support emotional well-being during language learning.

When negative emotions are detected, the system can deliver simplified, stress-reducing input, or adjust the tone of voice in audio feedback to provide a more calming, supportive experience.

Additionally, learners receive encouraging phrases and positive reinforcement to boost their self-confidence and engagement. This emotion-responsive approach creates a psychologically safe learning environment conducive to sustained motivation.

3.2 Multimodal Content for Learning Style Alignment

Learners possess diverse cognitive preferences—such as visual, auditory, or kinesthetic learning styles. MLLMs enable differentiated instruction by generating multimodal resources that align with these styles. For visual learners, diagram-based grammar explanations, image-annotated vocabulary, or infographics can be delivered. Auditory learners may benefit from AI-generated spoken dialogues, pronunciation drills, or listening comprehension tasks. Kinesthetic learners can engage in interactive role-playing scenarios or voice-activated simulations.

By tailoring the mode of content delivery, MLLMs enhance information processing efficiency, memory retention, and learner satisfaction. This alignment fosters greater accessibility and inclusiveness in language instruction, particularly in vocational contexts where engagement is critical.

3.3 Adaptive Feedback and Task Personalization

Traditional feedback mechanisms often fail to address learners' individual differences in real time. MLLMs overcome this limitation by analyzing students' language proficiency, response patterns, and error types to generate precise, personalized feedback. For instance, learners struggling with grammar may receive sentence-level corrections and focused explanations, while vocabulary-related issues may prompt semantic mapping or collocation exercises.

MLLMs also support dynamic task recommendation. Based on the learner's current stage within their Zone of Proximal Development (ZPD), the system automatically assigns practice activities of appropriate difficulty—such as vocabulary drills, speaking prompts, or reading comprehension tasks[9]. Teachers can customize task templates, while MLLMs adapt them to suit individual learner profiles. This strategy facilitates continuous progress and supports scaffolded skill acquisition.

3.4 Contextualized Output Task Generation

To enhance the practical application of second language skills, MLLMs can generate output tasks rooted in authentic, profession-specific contexts. By taking into account learners' academic majors, career aspirations, and interests, MLLMs produce communicative tasks that simulate real-world scenarios. For example, ceramic art majors may be given tasks like "introducing ceramic products at an international trade fair" or "writing a response to an English-speaking client's inquiry."

These contextualized tasks strengthen students' ability to use language in meaningful, industry-relevant settings, while also reinforcing their professional identity. MLLMs may further assist learners through collaborative writing, simulated dialogues, or vocabulary enhancement—bridging the gap between academic instruction and workplace communication demands.

4. Integrating Emotion Regulation and Vocational Relevance into MMLLM-Enhanced Pedagogy

4.1 Emotion Regulation Training

One of the most critical implications of this research is the need to incorporate explicit

instruction in emotion regulation within second language pedagogy. Many vocational undergraduates lack metacognitive awareness of their emotional states and possess limited strategies to manage negative emotions such as anxiety, frustration, or self-doubt. Embedding structured emotion regulation training into the curriculum—such as guided self-reflection, mindfulness exercises, and cognitive reappraisal techniques—can empower learners to take greater control of their affective experiences.

MMLLMs can support this training by modeling regulatory behaviors through interactive dialogue, scenario-based simulation, and real-time affective feedback [10]. For example, when a learner exhibits signs of cognitive overload, the system may offer calming prompts or reframe errors as growth opportunities. By reinforcing strategies like emotional labeling and self-monitoring through conversational interactions and feedback loops, MMLLMs serve not only as instructional agents but also as emotional mentors—helping students build resilience and self-efficacy over time.

4.2 Emotion-Aware Learning Design

Instructors must be equipped with the skills and frameworks to design emotionally responsive learning environments in collaboration with MMLLM technologies. This involves training teachers to interpret emotional cues—whether inferred through facial expressions, voice tone, or learner behavior—as indicators for pedagogical adjustment. MMLLMs can function as powerful co-facilitators in this process, capable of adapting instructional pacing, rephrasing task instructions, or altering the complexity of language input based on learners' emotional states[11].

Educators should therefore be supported in developing a dual literacy: technological fluency in navigating MMLLM interfaces and pedagogical sensitivity in leveraging AI-generated emotional insights for human-centered instruction. Professional development initiatives should emphasize affective scaffolding techniques, including the use of empathetic feedback, multimodal reassurance, and flexible task sequencing. Such an approach not only humanizes digital learning but also bridges the emotional disconnect often found in large or impersonal classroom settings.

4.3 Vocational Relevance in Content

Sustaining learner motivation and emotional engagement requires that instructional content be perceived as meaningful and professionally relevant. For vocational students whose language learning is closely tied to career aspirations, output tasks should be contextualized within their fields of study, such as tourism, international trade, or ceramic arts. MMLLMs are particularly well-suited for generating personalized, authentic, and industry-specific language scenarios that mirror real-world communication demands.

By simulating practical tasks—e.g., conducting a mock product pitch, writing an email to a foreign client, or participating in a virtual trade fair—MMLLMs provide learners with opportunities to use English in contextually grounded, emotionally resonant ways[12]. This not only aligns instruction with instrumental motivation but also supports the internalization of language use as a form of professional self-expression. When students perceive their L2 development as directly applicable to their future careers, they are more likely to invest sustained emotional and cognitive effort into learning.

5. Conclusion

This study highlights the centrality of emotion regulation in the second language (L2) learning experiences of vocational undergraduate students and foregrounds the transformative potential of multimodal large language models (MMLLMs) in addressing both the affective and cognitive

dimensions of L2 pedagogy. The emotional landscape of vocational learners is characterized by elevated language anxiety, diminished academic self-concept, a reliance on external motivation, and underdeveloped self-regulatory strategies. These emotional challenges pose significant barriers to sustained engagement and communicative competence, particularly in contexts where practical language use is critical for future professional success.

The integration of MMLLMs into language instruction offers a novel, technologically enhanced pathway for mitigating these barriers. By enabling real-time emotion-aware feedback, multimodal content adaptation aligned with individual learning styles, and contextually rich, profession-specific output tasks, MMLLMs reposition the language classroom as a dynamic, emotionally responsive, and learner-centered ecosystem. Rather than treating emotion regulation as an ancillary concern, the MMLLM-supported model embeds it as an integral component of instructional design—fostering emotional resilience, promoting learner autonomy, and enhancing motivation.

Ultimately, the study reaffirms that effective second language acquisition, particularly in vocational education settings, requires not only linguistic input and practice, but also sustained attention to the emotional experiences that shape learners' engagement and persistence. MMLLMs, when employed with pedagogical intentionality, have the capacity to cultivate affectively supportive learning environments that empower vocational students to overcome emotional constraints, actualize their communicative potential, and engage more confidently with real-world language tasks. These findings contribute to a growing body of research advocating for affect-sensitive, AI-augmented approaches to L2 education in applied and practice-oriented higher education contexts.

Acknowledgement

References

[1] Derakhshan, A., & Zare, J. (2023). The Impact of Altruistic Teaching on English as a Foreign Language (EFL) Learners' Emotion Regulation: An Intervention Study. Brain Sciences, 13(3), 458.

[2] Gross, J.J. (2015). Emotion regulation: Current status and future prospects. Psychol. Inq. 26, 1–26. [Google Scholar] [CrossRef]

[3] Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. Journal of personality and social psychology, 85(2), 348.

[4] Teng, L. S., & Zhang, L. J. (2018). Effects of motivational regulation strategies on writing performance: A mediation model of self-regulated learning of writing in English as a second/foreign language. Metacognition and Learning, 13(2), 213-240.

[5] Thompson, R. A. (1994). Emotion regulation: A theme in search of definition. Monographs of the society for research in child development, 25-52.

[6] Bielak, J., & Mystkowska-Wiertelak, A. (2020). Language learners' emotion-regulation strategies: A narrative review. Language Education and Emotions, 21-36.

[7] Webster, E. A., & Hadwin, A. F. (2015). Emotions and emotion regulation in undergraduate studying: Examining students' reports from a self-regulated learning perspective. Educational Psychology, 35(7), 794-818.

[8] De Neve, D., Bronstein, M. V., Leroy, A., Truyts, A., & Everaert, J. (2023). Emotion regulation in the classroom: A network approach to model relations among emotion regulation difficulties, engagement to learn, and relationships with peers and teachers. Journal of youth and adolescence, 52(2), 273-286.

[9] Kopelman-Rubin, D., Siegel, A., Weiss, N., & Kats-Gold, I. (2021). The relationship between emotion regulation, school belonging, and psychosocial difficulties among adolescents with specific learning disorder. Children & Schools, 42(4), 216-224.

[10] Zheng, S., & Zhou, X. (2022). Positive influence of cooperative learning and emotion regulation on EFL learners' foreign language enjoyment. International Journal of Environmental Research and Public Health, 19(19), 12604.

[11] Daros, A. R., Daniel, K. E., Meyer, M. J., Chow, P. I., Barnes, L. E., & Teachman, B. A. (2019). Impact of social anxiety and social context on college students' emotion regulation strategy use: An experience sampling study. Motivation and emotion, 43(5), 844-855.

[12] Rokni, S., Hamidi, H., & Ebadi, H. (2014). Relationship between emotional intelligence and emotion regulation of EFL learners. Int. J. Basic Sci. Appl. Res, 3, 376-380.