

# RESEARCH AND ADVANCES IN EDUCATION

---

ISSN 2788-7057

JUN. 2026 VOL. 5, NO. 2

---

## CONTENTS

**Fostering Responsible Conduct Among Trainees Through Integration of Ethics  
into the Curriculum: A Case Study of Nsamizi Training Institute of Social  
Development** **1-7**

Kikomeko Joseph, Muweesi Charles

**Generative AI-Enabled International Chinese Language Teaching: Innovations  
and Challenges** **8-13**

Ruyang Tang

**The Limited Classroom Application of Teacher Digital Literacy Training and Its  
Improvement Paths** **14-21**

Yuan Li

**Are We Really Listening? Development and Psychometric Validation of a Five-  
Stage Listening Process Scale for Higher Education Students in Region II,  
Philippines** **22-31**

Mark Kevin Astrero, Saranay I. Doyaoen, Darius King Galo, Aileen Melody De Vera, Titov Dela Cruz

**Discipline or Pressure? Examining the Effectiveness of the KDM Military  
Barracks in Developing Students' Learning Motivation** **32-39**

Desti Agustin

**The Influence of Demonstration Lessons on Teachers' Conceptions of Good  
Teaching** **40-47**

Hua Fang

# Fostering Responsible Conduct Among Trainees Through Integration of Ethics into the Curriculum: A Case Study of Nsamizi Training Institute of Social Development

Kikomeko Joseph<sup>1</sup> & Dr. Muweesi Charles<sup>2</sup>

<sup>1</sup> Muteesa I Royal University, Uganda

<sup>2</sup> Busitema University, Uganda

Correspondence: Kikomeko Joseph, Muteesa I Royal University, Uganda.

doi:10.63593/RAE.2788-7057.2026.06.001

## Abstract

Professional ethics constitutes a foundational aspect of vocational training; however, a considerable proportion of Ugandan curricula tends to emphasize technical skills while exhibiting inadequate consideration of moral development. This qualitative case study investigated the mechanisms by which the integration of ethics into the curriculum fosters responsible conduct among trainees at the Nsamizi Training Institute of Social Development, situated within Mpigi District. Guided by Kohlberg's Theory of Moral Development, the study employed an exploratory case study design. Data were collected through analysis of curriculum documents, semi-structured interviews with ten lecturers, and focus group discussions with twenty-five final-year students.

The findings suggested that a hybrid strategy, comprising both stand-alone ethics modules and their infusion into technical courses, may enhance trainees' cognitive apprehension of principles such as integrity, confidentiality, and professional responsibility. Nevertheless, the practical application of these principles appeared to remain inconsistent, potentially attributable to the pervasive influence of the hidden curriculum, extant societal corruption, and the paucity of practical ethical dilemmas presented. The investigation, therefore, posits that curriculum integration alone might prove insufficient without a congruent alignment between inculcated values and prevalent institutional practices. The recommendations encompass: (1) the establishment of an institutional Ethics Committee; (2) the incorporation of practicum-based ethics assessments; (3) the provision of faculty training in ethics pedagogy; and (4) governmental enforcement of ethical codes for interns and practitioners, all of which endeavor to bolster ethical training within social development institutions.

**Keywords:** curriculum integration, professional ethics, responsible conduct, moral development, Nsamizi Training Institute, vocational training, hidden curriculum

## 1. Introduction

The integration of ethics into education traces back to ancient philosophical traditions, notably Aristotle's concept of *ethos*, which emphasized character formation through habituation and practical wisdom. In professional and vocational training, formalized ethics education emerged prominently in the 20th century alongside the growth of professional associations that established codes of conduct in fields such as medicine, engineering, law, and social work (Sulmasy, 2020). Globally, concerns over professional misconduct, ranging from corruption to negligence, prompted pedagogical shifts toward embedding ethics within disciplinary curricula rather than treating it as an isolated subject (Fort, 2021).

In Uganda, the history of ethics in higher and vocational education is intertwined with the liberalization of the education sector in the 1990s, which expanded access but also raised questions about quality and moral

standards (2022). Early approaches often confined ethics to standalone “General Studies” or humanities courses. Over the past two decades, however, there has been a gradual move toward infusion models, where ethical concepts are embedded within technical or professional modules to enhance relevance and application (Kuteesa, 2021). This evolution reflects broader national development priorities. Uganda’s Vision 2040 positions ethical humanism and moral responsibility as foundational pillars for sustainable socio-economic transformation (Muyingo, 2018).

Nsamizi Training Institute of Social Development (NTISD), founded in 1952/1953 in Mpigi District as a center for training social welfare and community development workers, exemplifies this context. The institute has historically focused on producing professionals equipped to address structural social challenges such as poverty, child protection, gender-based violence, and community empowerment in both urban and rural settings (Twikirize, 2014). Graduates are expected to uphold high standards of integrity, confidentiality, and social justice, particularly when working with vulnerable populations through government departments, NGOs, and community-based organizations.

Despite these expectations, current issues reveal a persistent gap. Ugandan society continues to grapple with widespread corruption, negligence, and unprofessional conduct among technically skilled graduates, including those in social development fields. Reports highlight cases where practitioners engage in bribery, falsification of records, or exploitation of clients, undermining public trust and development efforts (Muwanga, 2018). In the education sector itself, phenomena such as the “hidden curriculum”, where institutional practices contradict formal teachings on honesty, exacerbate the problem (Musenze, 2021). Large class sizes, overloaded curricula, and limited faculty preparation for ethics instruction further compound these challenges (Bennaars, 2019; Ejuu, 2022).

This study is important because technically competent but ethically deficient professionals pose risks to vulnerable communities and national development goals. By focusing on Nsamizi, a premier public tertiary institution dedicated to social development training, the research illuminates practical strategies and barriers within a real-world vocational context.

### *1.1 Justification of Key Concepts*

*Responsible Conduct* denotes behavior aligned with professional codes, moral principles, and societal expectations. It represents the behavioral outcome of effective ethics education rather than mere knowledge acquisition (Namukasa, 2023). *Curriculum Integration* involves systematically embedding ethical concepts into technical modules instead of isolating them, justified by evidence that context-specific learning enhances relevance and retention (Bakabulindi, 2020; Kuteesa, 2021). *Trainees* refer to adult learners at Nsamizi, who enter programs with pre-existing value systems shaped by cultural, socio-economic, and personal experiences. Unlike schoolchildren, their moral reasoning requires critical engagement to challenge and advance existing frameworks (Kohlberg, 1981). These concepts are central to understanding how Nsamizi can bridge the persistent “knowing-doing” gap in professional ethics.

### *1.2 Problem Statement*

What is already known is that ethics education improves cognitive moral reasoning and that integrated approaches are generally more effective than isolated courses in professional training (Fort, 2021; Rest, 1986; Sulmasy, 2020). In Uganda, studies document high awareness of ethical principles among students but frequent justification of corrupt practices as “survival” mechanisms (Muwanga, 2018). Vocational and social work training institutions, including those like Nsamizi, have introduced ethics components in response to national concerns over graduate misconduct.

The problem, however, is evident in the rising cases of unethical conduct among social development practitioners, such as misappropriation of funds, breach of confidentiality, and exploitation, despite technical training. This misconduct erodes public trust, hampers service delivery to vulnerable populations, and contradicts Uganda’s development aspirations (Opolot, 2019). Evidence from the education sector shows that institutional hypocrisy and the hidden curriculum often undermine formal ethics teaching (Ejuu, 2022; Musenze, 2021).

Past studies have largely focused on standalone ethics courses or general higher education, with limited attention to infusion models in vocational social development training. Few have examined real institutional cases like Nsamizi or explored the interplay between curriculum strategies, trainee application, and contextual challenges in Uganda’s corruption-prone environment (research gap).

This research must be conducted to provide evidence-based insights into effective integration strategies while identifying practical barriers. The proposed solution is a comprehensive examination of Nsamizi’s mixed infusion approach, offering actionable recommendations to align curriculum, institutional culture, and societal support for fostering genuine responsible conduct.

### *1.3 Purpose of the Paper*

This investigation seeks to elucidate the mechanisms through which the integration of ethical principles into the curriculum may foster responsible conduct among trainees at the Nsamizi Training Institute of Social Development.

### *1.4 Research Objectives*

The present investigation was delineated by the subsequent objectives:

- 1) To delineate the pedagogical strategies employed for the integration of ethical principles within the curriculum at the Nsamizi Training Institute.
- 2) To assess the comprehension and practical application of ethical principles by trainees during their programmatic instruction.
- 3) To evaluate the specific challenges encountered by instructional staff in promoting responsible conduct via curriculum integration.

### *1.5 Research Questions*

This investigation was designed to address the subsequent interrogatives:

- 1) What pedagogical strategies are currently employed for the integration of ethical principles within the curriculum at the Nsamizi Training Institute?
- 2) In what manner do trainees comprehend and practically apply ethical principles during their programmatic instruction?
- 3) What specific challenges do instructional staff encounter in fostering responsible conduct through the integration of curriculum components?

### *1.6 Significance of the Study*

- 1) **Academic Contribution:** This investigation is anticipated to augment the existing body of knowledge pertaining to ethics pedagogy within vocational and social work education in sub-Saharan Africa, thereby furnishing empirical insights for curriculum developers and scholarly researchers.
- 2) **Professional Practice and Industry:** Social development practitioners and instructional personnel at institutions such as Nsamizi may acquire tangible strategies for ameliorating the theory-practice disparity, thereby enhancing professional integrity in practical field applications.
- 3) **Policy Formulation:** The emergent findings are projected to inform the National Council for Higher Education (NCHE), the Ministry of Education and Sports, and the Ministry of Public Service concerning the development of policies for ethical integration and the rigorous enforcement of professional codes of conduct.
- 4) **Societal Impact:** Through the promotion of responsible conduct among prospective social workers, this investigation may facilitate enhanced service provision to vulnerable communities, a diminution of corruption, and the strengthening of social cohesion within Uganda.

### *1.7 Scope of the Study*

- 1) The delimitation of this investigation pertains to the Nsamizi Training Institute of Social Development (NTISD) in Mpigi District, Uganda; it serves as a single-site case study designed to offer an in-depth analysis of ethics education. The investigative focus encompasses final-year trainees, lecturers, and curriculum developers within social work and development programs, as these stakeholders offer the most relevant insights into the efficacy of current pedagogical strategies.
- 2) Regarding its substantive scope, this investigation is centered upon three pivotal areas: the strategies employed for curriculum integration, the level of trainees' ethical understanding and application, and the challenges that impede the fostering of responsible conduct. The investigation covers the 2023–2025 academic period, thus enabling an assessment of the institute's most current practices.
- 3) Although the qualitative design provides rich, contextual data, it inherently limits the generalizability of the findings beyond NTISD. Furthermore, the reliance upon participant self-reporting may present a potential for social desirability bias. These limitations were addressed through the rigorous triangulation of data sources, which included document analysis and focus groups, thereby verifying accuracy and strengthening the validity of the conclusions.

### *1.8 Theorization*

The theoretical underpinning for this investigation is Lawrence Kohlberg's Theory of Moral Development (Kohlberg, 1981). Kohlberg posited that moral reasoning progresses through three distinct levels, namely,

Pre-conventional (characterized by a focus on punishment/reward and self-interest), Conventional (marked by conformity to social norms and authority), and Post-conventional (defined by adherence to universal ethical principles and social contracts). Each of these levels encompasses two stages, with advancement potentially occurring through cognitive conflict and exposure to higher-stage reasoning via moral dilemmas and dialectical engagement.

The theory explains moral development as a cognitive process stimulated by an active engagement with moral issues, as opposed to passive adherence to rules. The theory has been widely applied within professional ethics education for the design of interventions that foster progression toward principled reasoning (McLeod, n.d.; structural-learning.com, 2023).

Kohlberg's theory is chosen because it directly addresses the study's focus on moving trainees beyond rule compliance toward internalized professional ethics. It justifies active integration methods, such as case studies, reflective journals, and ethical dilemmas, over didactic teaching. In the context of Nsamizi, the theory relates to the topic by framing curriculum integration as a mechanism for creating cognitive disequilibrium that advances trainees from conventional (societal approval or fear of punishment) to post-conventional reasoning, where they uphold integrity even in corrupt environments. This supports the examination of strategies that stimulate moral growth and the challenges that hinder progression, such as contradictory institutional practices. A supplementary lens from Rest's (2020) Four Component Model (sensitivity, judgment, motivation, character) complements Kohlberg by emphasizing the behavioral application gap observed in findings.

## **2. Literature Review**

### **Variable 1: Strategies for Curriculum Integration**

Background and global studies show varied approaches, including case-based learning in medicine (Sulmasy, 2020), infusion in business (Fort, 2021), and hybrid models in TVET (Kuteesa, 2021). In Africa, service learning and role-plays connect ethics to community realities (Nakkazi, 2022). Faculty training and team-teaching are critical, as technical instructors often lack ethics preparation (Bennaars, 2019). What previous studies found is that context-specific, experiential strategies enhance relevance and retention. What is missing is in-depth case analysis of infusion in Ugandan social development vocational institutes, particularly how local corruption examples function as teaching tools.

### **Variable 2: Trainees' Understanding and Application**

Understanding is cognitive, while application is behavioral. Rest's (2020) model highlights the need for sensitivity, judgment, motivation, and character. Ugandan studies reveal that students define ethical terms well but justify misconduct due to survival pressures (Muwanga, 2018). Field-based training improves integrity during internships (Nabwire, 2023). Previous findings indicate assessment tied to conduct strengthens the application (Opolot, 2019). Missing is an exploration of how workplace role models mediate curriculum effects in social development training.

### **Variable 3: Challenges in Fostering Responsible Conduct**

The hidden curriculum often contradicts formal teaching (Musenze, 2021). Overloaded curricula, subjective assessment, cultural relativism, and large classes pose barriers (Ejuu, 2022; Bakabulindi, 2020; Namukasa, 2023). In Uganda, institutional hypocrisy in education undermines ethics efforts. Studies elsewhere note similar disjunctions in responsible management education (Olanya, 2023) issuing is institution-specific analysis of how these challenges manifest at vocational social work institutes and their impact on moral development progression per Kohlberg.

## **3. Methodology**

The study adopted a qualitative exploratory case study design to gain an in-depth understanding of ethics integration at a single institution. The population comprised curriculum developers, lecturers, and final-year trainees at Nsamizi Training Institute. Purposive sampling selected 10 lecturers experienced in ethics-infused courses, while stratified random sampling selected 25 final-year students from social work and development programs.

Data collection involved semi-structured interviews (guided by the three research questions, with four probe questions each), curriculum document analysis (syllabi and course outlines), and focus group discussions. Thematic analysis was employed to identify patterns, with codes developed around strategies, understanding/application, and challenges. Trustworthiness was ensured through triangulation, member checking, and an audit trail.

Table 1. Distribution of Study Respondents

| Category     | Role                                 | Number    | Rationale for Selection                             |
|--------------|--------------------------------------|-----------|-----------------------------------------------------|
| Lecturers    | Course tutors / Department heads     | 10        | Experienced in delivering ethics-infused courses    |
| Students     | Final-year Social Work & Development | 25        | Have completed ethics modules and field attachments |
| <b>Total</b> |                                      | <b>35</b> |                                                     |

The table shows that 35 respondents participated. Lecturers offered insights into curriculum design and teaching challenges, while students provided perspectives on understanding, application, and lived experiences. This distribution enabled rich, triangulated qualitative data while remaining manageable for in-depth analysis.

#### 4. Interview Findings

##### Finding on Objective 1: Strategies for Integration

Nsamizi employs a mixed “stand-alone plus infusion” model. Technical modules dedicate sessions to ethics (e.g., “honesty in budgeting” in project planning). Lecturers use local newspaper cases of NGO fund misappropriation and require reflective journals during field attachments. This aligns with hybrid approaches recommended in Ugandan TVET (Kuteesa, 2021) but adapts them to social development realities.

The use of real Ugandan examples makes ethics concrete and relevant, bridging theory-practice gaps as per Kohlberg’s emphasis on cognitive conflict. However, implementation varies by lecturer preparedness, suggesting the need for standardized guidelines to ensure consistency across departments.

##### Finding on Objective 2: Trainees’ Understanding and Application

Trainees demonstrated a strong cognitive grasp of principles like integrity and confidentiality but reported inconsistent application. Some “cut corners” when supervisors modeled unethical behavior, yet others cited classroom debates as giving them the courage to refuse to falsify records. This mirrors the knowing-doing gap in Ugandan higher education (Muwanga, 2018; Nabwire, 2023).

Curriculum integration builds moral awareness and initial courage, supporting progression toward post-conventional reasoning. Nevertheless, environmental influences heavily mediate actual behavior, highlighting that cognitive gains alone do not guarantee responsible conduct without supportive field contexts.

##### Finding on Objective 3: Challenges Faced by Lecturers

Prominent challenges included institutional hypocrisy (e.g., bribes for administrative services), large classes hindering dialogue, outdated materials, student resistance to “preachy” topics, and subjective assessment of ethics. The hidden curriculum was frequently cited as undermining teachings. These echo broader Ugandan and African education challenges (Ejuu, 2022; Musenze, 2021).

Such contradictions create cognitive dissonance for both lecturers and trainees, stalling moral development at conventional levels. Addressing these requires systemic alignment between formal curriculum and institutional practices to make ethics education credible and effective.

#### 5. Discussion

Discussion on strategies aligns findings with global and regional best practices while noting contextual adaptations at Nsamizi. Infusion combined with local cases enhances relevance (Sulmasy, 2020; Fort, 2021; Kuteesa, 2021), and reflective journals promote metacognition central to Kohlberg (1981). However, variability in implementation points to the need for faculty development (Bennaars, 2019).

Discussion on understanding and application confirms cognitive gains but persistent behavioral gaps influenced by role models and environment (Rest, 2020; Muwanga, 2018; Nabwire, 2023). This supports Kohlberg’s stage progression model while underscoring external variables.

Discussion on challenges validates the destructive role of the hidden curriculum and societal factors (Musenze [CITE\_1]; Ejuu [CITE\_2]; Olanya [CITE\_3]). Without cultural and institutional alignment, integration efforts may risk failure, as cognitive conflict alone cannot overcome contradictory modeling.

#### 6. Conclusions and Recommendations

##### 6.1 Conclusions

Integrating ethics into the curriculum at Nsamizi significantly enhances trainees’ cognitive understanding of

professional responsibility and provides tools for moral courage. Nevertheless, the translation into consistent responsible conduct is hindered by the hidden curriculum, societal corruption pressures, and insufficient practical application opportunities. Kohlberg's framework elucidates that while curriculum strategies stimulate reasoning, progression to post-conventional levels likely necessitates a supportive ethical ecosystem. The veritable fostering of responsible conduct ostensibly demands a congruence between taught values, institutional practices, and broader societal norms. Without this alignment, vocational training in social development might risk producing graduates who comprehend ethics yet subsequently fail to enact them, thereby potentially perpetuating cycles of misconduct.

## 6.2 Recommendations

- 1) Institutions such as Nsamizi should establish a standing Ethics Committee to monitor internal operations for consistency with taught values, including transparent assessments and anti-corruption measures.
- 2) Curriculum developers, in collaboration with the National Council for Higher Education (NCHE), should strengthen practicum ethics assessments by incorporating supervisor feedback on observed behavior.
- 3) Lecturers, for the purpose of fostering ethical discussions within expansive classes and ensuring the objective assessment of moral growth, demonstrably necessitate regular capacity-building workshops.
- 4) Through the Ministry of Public Service and the Ministry of Education, the Government should endeavor to enforce a uniform code of ethics for all interns and social development practitioners, an initiative which ought to be coupled with robust monitoring mechanisms.
- 5) Finally, educationists and professional bodies like the National Association of Social Workers of Uganda (NASWU) should promote ongoing research and sharing of best practices in ethics infusion to build a national culture of responsible professional conduct.

## References

- Bakabulindi, G. (2020). Assessing moral character in vocational training. *Journal of Vocational Education & Training*, 72(4), 550-565. <https://doi.org/10.1080/13636820.2020.1781689>
- Bennaars, G. (2019). *Philosophy of education in Uganda*. JKP.
- Ejuu, G. (2022). Education for sustainable development in Uganda: The ethical dimension. *International Journal of Educational Development*, 88, Article 102585. <https://doi.org/10.1016/j.ijedudev.2022.102585>
- Fort, T. (2021). Ethics in business schools: A literature review. *Journal of Business Ethics*, 167(2), 231-247. <https://doi.org/10.1007/s10551-019-04270-6>
- Kohlberg, L. (1981). *The philosophy of moral development*. Harper & Row.
- Kuteesa, S. (2021). Infusion or isolation? Strategies for teaching ethics in Ugandan universities. *African Journal of Higher Education*, 9(2), 78-95. <https://doi.org/10.4314/ajhe.v9i2.3>
- Musenze, I. (2021). The hidden curriculum and its impact on student moral development. *Uganda Journal of Education*, 7(1), 112-128.
- Muwanga, D. (2018). Corruption and the moral reasoning of university students in Kampala. *Journal of African Studies*, 12(1), 45-60. <https://doi.org/10.1017/S000197201800012X>
- Muyingo, J. (2018). *Technical education and human resource development in Uganda*. Fountain Publishers.
- Nabwire, S. (2023). Field work and professional integrity among social work students. *Social Work Education*, 42(3), 345-360. <https://doi.org/10.1080/02615479.2023.2172884>
- Namukasa, D. (2023). Responsible conduct in the digital age: Challenges for trainees. *East African Journal of Information Technology*, 5(1), 22-38.
- National Council for Higher Education. (2022). *Guidelines for quality assurance in higher education*. NCHE.
- Opolot, D. (2019). The dilemma of teaching ethics in a corrupt society. *Ethics and Social Welfare*, 13(2), 150-165. <https://doi.org/10.1080/17496535.2019.1579865>
- Rest, J. (2020). *Moral development: Advances in research and theory*. Praeger.
- Sulmasy, D. (2020). A brief history of medical ethics. *AMA Journal of Ethics*, 22(4), E344-E350. <https://doi.org/10.1001/amajethics.2020.344>
- Twikirize, J. M. (2014). Social work education and practice in Uganda: Challenges and prospects. In *Social work education and practice in Africa* (pp. 139-156).

**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

# Generative AI-Enabled International Chinese Language Teaching: Innovations and Challenges

Ruyang Tang<sup>1</sup>

<sup>1</sup> School of Languages and Cultures (School of International Communication and Exchange), Shanghai University of Political Science and Law, Shanghai 201701, China

Correspondence: Ruyang Tang, School of Languages and Cultures (School of International Communication and Exchange), Shanghai University of Political Science and Law, Shanghai 201701, China.

doi:10.63593/RAE.2788-7057.2026.06.002

## Abstract

The rapid advancement of generative artificial intelligence has reshaped traditional operational systems across numerous social sectors. In China, increasing emphasis has been placed on educational applications of AI, with consistent support for exploring the potential of generative models in teaching and learning. In 2025, the Ministry of Education issued the Digital Strategy for Education, which promotes the construction of new educational infrastructure and encourages universities to develop innovative “AI + Education” models. Through deep learning and algorithm-driven pattern simulation, generative AI can support teachers in lesson planning, resource creation, and AI-integrated textbook development. It also facilitates personalized learning and helps cultivate talents with strong information literacy. Supported by AI technologies, the “HI + AI” teaching framework creates a tripartite collaborative system involving teachers, students, and machines, leading to more adaptive and efficient classroom practices. While generative AI opens new prospects for international Chinese language education, it also brings emerging challenges that require systematic reflection and response.

**Keywords:** artificial intelligence in education, international Chinese language teaching, generative artificial intelligence

## 1. Introduction

The rapid development of artificial intelligence has reshaped the traditional ecosystems of various sectors in society. As reported in the 2023 government work statement, China will continue to advance the “AI+” initiative and support the widespread application of large-scale models, marking a shift from the “Internet+” era to the “AI+” era. AI has also profoundly impacted the field of education. In late 2022, OpenAI launched the ChatGPT-3.5 model, positioning generative AI at the forefront of AI technologies. On January 15, 2025, the DeepSeek application was officially released, once again sparking reflection on the potential applications of generative AI in education. The Chinese government places a high priority on AI in education. In January 2025, the Central Committee of the Communist Party of China and the State Council issued the Outline for the Construction of a Leading Country in Education, which states the need to “promote AI-driven educational transformation.” The outline emphasizes leveraging AI to advance educational equity and personalized development in China and to facilitate digital transformation. The 2025 Digital Strategy for Education issued by the Ministry of Education actively promotes the construction of new educational infrastructure, encouraging colleges and universities to explore novel models of “AI + Education.” As international Chinese language education constitutes an important component of the broader educational enterprise, how to harness AI to empower international Chinese language teaching has become a question of ongoing reflection and practice for frontline teachers.

At present, the field of international Chinese language education has seen considerable discussion regarding the

use of AI, including its advantages and disadvantages. Beyond supporting instructional processes, AI can also provide multilingual outreach and customized services in other areas, such as student admissions (Du & Zheng, 2025). However, existing research outcomes remain predominantly focused on macro-level frameworks, future application scenarios, and general principles, with a relative lack of empirical research and practice-based instructional analyses. In response, this study adopts a practice-oriented research approach to explore the concrete application scenarios of generative AI technologies in international Chinese language education. It also aims to synthesize the specific difficulties encountered by teachers and students in the application process and to propose corresponding solutions, thereby exploring new pathways for the digital transformation of international Chinese language education.

## 2. Literature Review

To date, scholars both domestically and internationally have conducted extensive research on the application of generative artificial intelligence (AI) in education, yielding fruitful results. Research topics span various levels of technological application, including the use of generative AI for curriculum design, instructional assessment, and analysis of student learning behaviors. Some scholars have carried out empirical studies examining the impact of generative AI on students' learning interest and academic performance. Regarding the integration of generative AI into language teaching, research has primarily focused on the following directions: generative AI-assisted language skills training, adaptive learning and personalized instruction, teaching resource development, and evaluation of teaching effectiveness. Scholars generally acknowledge the value of generative AI in developing instructional resources, including smart hardware, smart software, and latent intelligent systems (Ma & Xu, 2025). Generative AI can facilitate personalized and individualized student learning, and human-machine collaborative learning approaches can significantly enhance learning efficiency. For teachers, generative AI offers assistance in designing lesson plans, generating exercises, and grading assignments. For students, it serves as a 24/7 online teaching assistant, providing customized learning resources and pathways.

The rapid advancement of generative AI also presents challenges for teachers, primarily concerning how to enhance their digital literacy to better utilize generative AI technologies, as well as how to improve relevant policies and regulations to ensure the security and standardization of technological applications. Ethical issues arising from the use of AI in teaching, such as data privacy protection and algorithmic bias, have also been the subject of in-depth discussion.

Nevertheless, existing research has yet to thoroughly investigate the specific application models of generative AI in international Chinese language teaching. There remains a lack of empirical research, practice-oriented exploration, and studies on long-term effects. The integration of generative AI with international Chinese language teaching stands at a crossroads where opportunities and challenges coexist. Future development should pursue the parallel paths of “technological integration” and “ethical governance,” develop “AI-teacher” collaboration models, and provide intelligent support for education under the principles of human-centeredness and human-machine collaboration.

## 3. Feasibility Analysis of Integrating Generative AI into International Chinese Language Teaching

### 3.1 Technical Concepts and Principles

Generative artificial intelligence is an important branch of AI. Unlike traditional discriminative AI, whose core function lies in analyzing and judging existing data, generative AI is characterized by its ability to generate novel content. Through deep learning and the simulation of inherent patterns in phenomena, and based on algorithms, models, and rules, generative AI can autonomously create new content, which can be output in forms such as text, images, audio, video, and code.

The technical principles of generative AI primarily rely on technologies such as deep learning and natural language processing (NLP). By constructing neural network models with multiple layers, computers automatically learn features and patterns from vast amounts of text data. Commonly used deep learning models in generative AI include generative adversarial networks (GANs), variational autoencoders (VAEs), and large language models (LLMs) based on the Transformer architecture. Generative AI possesses robust capabilities in natural language understanding and generation. When a user inputs a prompt, the model generates coherent text, images, or other forms of content based on learned knowledge and probabilistic computations, while supervised fine-tuning (SFT) and reinforcement learning from human feedback (RLHF) enable controllability of the generated output. Mainstream LLMs typically exhibit the following characteristics: first, they have massive parameter sizes, enabling them to learn rich linguistic knowledge and semantic information; second, they possess strong generalization capabilities, performing well across a variety of NLP tasks; and third, through large-scale unsupervised learning and supervised fine-tuning, they continuously optimize model performance and effectiveness.

Based on the above characteristics, generative AI holds immense potential for application in educational contexts.

It enables on-demand synthesis of personalized learning materials, multi-dimensional analysis and reconstruction of complex concepts, and the provision of dynamic cognitive feedback through conversational interaction.

### *3.2 The Need for Talent Development*

On one hand, the widespread societal application of AI is a foreseeable future trend. In this context, AI literacy, as a core dimension of digital literacy, is gradually becoming a key component of the competency structure of modern talent. How to evaluate the technical characteristics of different AI platforms based on task requirements and make optimal choices, how to issue instructions and interact with AI logic to achieve one's goals, and how to select and critically evaluate the vast amount of information provided by generative AI have all become new "required competencies." This underscores the necessity of introducing generative AI into the classroom to cultivate students' professional capabilities and guide them in adapting to societal demands.

On the other hand, in the era of AI, global educational objectives are undergoing a paradigm shift. The international community's criteria for evaluating talent competencies have expanded from a single dimension of specialized skills to a comprehensive competency framework encompassing cognitive abilities, expressive capabilities, and creativity. This shift also imposes new requirements on international Chinese language teaching. Integrating generative AI into international Chinese language instruction can help cultivate well-rounded individuals with positive affect, correct attitudes, and holistic intellectual development.

Specifically, the development of cognitive abilities includes the monitoring and regulation of one's own cognitive processes, as well as the understanding and adaptation to multicultural societies. This requires learners to continuously optimize their learning strategies through reflective practice, particularly in achieving knowledge transfer across intercultural contexts. Expressive capabilities must go beyond linguistic mastery to emphasize communicative adaptability grounded in cultural backgrounds—that is, the ability to dynamically adjust discourse strategies according to different cultural contexts. This places higher demands on the design of situated teaching activities in international Chinese language instruction. As generative AI gradually replaces procedural work tasks, the cultivation of creativity becomes the core value of education. In essence, creativity involves fostering learners' critical thinking and innovative problem-solving abilities that transcend established patterns.

### *3.3 The Need for the Development of International Chinese Language Teaching Itself*

In 2024, UNESCO released AI literacy frameworks for students and teachers, respectively. These frameworks require teachers to use AI tools correctly and creatively in teaching, based on a human-centered approach. In April 2025, at the World Internet Conference, discussions on educational transformation in the AI era proposed that education should integrate generative AI with human intelligence to enhance educational quality. In May 2025, the white paper on Smart Education in China, released by the Ministry of Education of the People's Republic of China, comprehensively introduced China's philosophies, approaches, initiatives, and achievements in advancing educational digitalization and developing smart education, sharing practical experiences in educational development and transformation in the intelligent era.

AI-empowered education is a challenge that all academic disciplines will face in their future development. In recent years, international Chinese language teaching has been undergoing a critical transition from "specialization" toward "popularization, universal access, and application-orientation." This process confronts three key challenges. First, the number of Chinese language learners worldwide is increasing, yet there is a shortage of high-quality teachers, resulting in a severely imbalanced student-teacher ratio. Second, there is a contradiction between traditional teaching models and the diverse, personalized learning needs of students. Overseas Chinese language learners are trending younger and toward non-degree pursuits, with significantly increased demand for engaging and personalized learning experiences. Third, it is difficult to balance depth of cultural transmission with teaching efficiency. Existing teaching resources and methods struggle to achieve the coordinated development of language proficiency and cultural cognition within limited instructional hours. The integration of generative AI offers new possibilities for addressing these contradictions. Generative AI can generate personalized linguistic materials on demand, achieving a revolutionary breakthrough in the supply of teaching resources, and can promote a shift in classroom ecology from "teacher-centered" to "human-machine collaborative," with the teacher's role evolving toward that of a "classroom designer." AI-based language practice systems can transcend spatiotemporal constraints, providing students with on-demand answers to their questions at any time and place, while significantly increasing the volume of language input and output. Using AI as an assistive tool to empower international Chinese language teaching can better construct an educational ecosystem of "AI-human intelligence collaboration," achieve inclusive innovation, and promote educational equity and sustainable development.

## **4. Pathway Analysis of Generative AI-Enabled International Chinese Language Teaching**

In a survey on "AI Usage" conducted with the author's participation, among 130 participants, approximately

16.2% used AI for teaching assistance, and 23.8% employed structured prompt frameworks. Only 6.9% had never used AI, and a very small minority (1.5%) reported a poor AI experience. Regarding participants' feedback and expectations for AI applications in teaching, several representative comments emerged: "AI has broad applications in teaching," "I hope to use AI to learn PPT/video creation to enrich the classroom," and "I look forward to AI teaching assistant functions." Some participants also reported that they "do not know how to use it" or "need to learn," expressing the hope that AI platforms would provide basic operation guides for beginners to alleviate anxiety about "being left behind." Generative AI has wide-ranging applications in international Chinese language teaching, which can be concretely examined from the following three aspects.

#### *4.1 Generation of Teaching Resources*

##### *4.1.1 Assisting Teachers in Lesson Preparation*

Assisting in generating lesson plans. By providing specific teaching content and a sample lesson plan to Kimi and inputting the instruction, "Based on the content of Document XX, design a lesson for students of XX grade using an XX-oriented teaching approach. The project theme is XX. The design of the teaching plan should follow the format of the second document," Kimi can generate a detailed lesson plan that includes specific teaching content, instructional stages, project activities, and teaching evaluations.

Generative AI can generate text, images, audio, video, and other resources according to teachers' needs. For example, if a teacher needs to design a lesson on the traditional Chinese festival of the Dragon Boat Festival, they need only input relevant themes and requirements into Deepseek—such as vocabulary difficulty at the HSK Level 4, key points including the origin and customs of the festival—and the system can quickly generate a passage that meets the requirements. The generated passage not only contains accurate linguistic expression and vivid descriptions but also incorporates elements of Chinese culture, including historical stories, culinary culture, and folk customs, making it both informative and engaging. For instance, "The taste of zongzi in the south and north is somewhat different; southerners prefer savory zongzi, while northerners prefer sweet zongzi." A vocabulary list is also provided at the end of the passage. When teaching vocabulary such as "dragon boat," "sachet," and "zongzi," visual aids such as images can be used. Teachers can use Doubao to generate images by inputting the instruction, "Generate a Dragon Boat Festival-themed image showing adults and children eating zongzi while watching a dragon boat race, with sachets hanging on their bodies." Finally, teachers can use Dujia or Sora to convert text into video with one click, saving considerable time that would otherwise be spent searching for suitable video materials, and enabling students to experience the Dragon Boat Festival atmosphere more intuitively. After class, teachers can find extended reading materials online and send the article link to the "AI Podcast" feature in Doubao, which will automatically generate a voice podcast of the article with standard pronunciation and natural, fluent human speech.

##### *4.1.2 Promoting Personalized Learning: Developing AI-Integrated Textbooks*

Existing textbooks have fixed content and cannot meet the personalized needs of students with varying learning speeds within the same class. Future textbook development should emphasize personalization and interactivity. With the help of generative AI, on the basis of content appropriate for the average student, additional extended content can be included. For example, more difficult exercises or extended reading materials can be converted into QR codes printed in the textbook, allowing advanced students to scan and learn on their own without feeling idle in class. This approach both accommodates the constraints of printed textbook content and meets the learning needs of more capable students.

University education should prepare students for their future career development. The "Chinese + Vocational Skills" education model has attracted widespread attention since its introduction, as it helps both international Chinese education and vocational education reach a global audience. However, currently, the development of "Chinese + Vocational" textbooks remains limited; their professionalism and suitability fall short of demand. Moreover, as society evolves rapidly, the content of older textbooks lags behind contemporary realities. Generative AI, with its powerful capabilities in data collection, analysis, and processing, can generate corresponding textbook content based on instructions. Taking Business Chinese as an example, to cultivate students' awareness of international business strategy and practical skills, as well as to deepen their understanding of China's business cultural environment, universities may offer courses such as "Cross-Border E-Commerce Practices" and "Business Etiquette." When suitable textbooks are unavailable, teachers can use generative AI to generate customized textbooks according to students' language proficiency levels and program-specific training objectives. Teachers may even adjust textbook content in real time in response to dynamic social changes to meet learning needs.

#### *4.2 Enrichment of Teaching Models*

##### *4.2.1 The "HI + AI" Teaching Model*

At present, leading universities in China have gained substantial experience in exploring the "HI + AI" teaching

model. HI (Human Intelligence) refers to human intelligence, specifically the guidance provided by teachers in the classroom, while AI refers to the various AI-based learning platforms selected by teachers. In Bloom's taxonomy of educational objectives, the cognitive domain is subdivided into six levels: memory, comprehension, application, analysis, evaluation, and creation. In the "HI + AI" teaching model, teachers integrate knowledge at the "memory and comprehension" levels into AI-mediated instruction, allowing students to engage in learner-led adaptive learning. Subsequently, training in skills at the "application, analysis, evaluation, and creation" levels becomes the focus of HI-mediated instruction, where teachers guide students in achieving higher-order learning objectives through methods such as case studies, hands-on practice, and project-based learning. This teaching model, empowered by AI, achieves a triadic collaboration among "teacher, student, and machine," making the classroom more adaptive and efficient. While enabling personalized learning, it also facilitates group collaboration, stimulates students' learning interest, and increases classroom participation.

#### 4.2.2 Contextualized Teaching

By creating authentic linguistic contexts for students in the classroom, teachers enable students to learn and apply knowledge in context. This approach not only improves students' Chinese expression skills and increases learning engagement but also exercises their use of Chinese communicative strategies and promotes the construction of practical knowledge. Situated teaching with generative AI participation transcends previous limitations. It is not constrained by existing human knowledge frameworks; it can customize content for specific scenarios such as job interviews or business contexts, enable personalized, tiered instruction, allow repeated practice unrestricted by spatiotemporal constraints, and provide immediate feedback and data-based assessment. Taking an advanced Chinese oral course as an example, after students have mastered foundational knowledge such as vocabulary, grammar, and sentence patterns at the memory and comprehension levels, teachers can use Deepseek or Doubao to conduct AI-simulated dialogue practice. For instance, by inputting the instruction, "I am an international student studying in China. My Chinese proficiency is HSK Level 5, and I want to find a job in China. Please pretend to be an interviewer and ask me questions, simulating the interview process," students can begin practicing in an immersive context. The AI can adjust speech rate and vocabulary complexity according to the student's Chinese proficiency level (e.g., HSK Level 5), gradually introduce specialized vocabulary, and provide suggestions for improvement based on the student's responses. AI-simulated interviews eliminate the anxiety of real-life scenarios, allowing students to build confidence through repeated practice, and enhance their cross-cultural communicative competence. The AI's immediate feedback can identify cultural offenses in students' communication, and tasks such as analyzing differences in consumption habits between China and Russia can cultivate students' cultural comparison skills, facilitating their progression from "literal translation responses" to "culturally adapted responses" in interviews.

#### 4.3 Innovation in Assessment Methods

Generative AI can automate assignment grading and learning outcome assessment while providing targeted feedback to students, greatly improving teaching efficiency and assessment accuracy. In terms of learning outcome assessment, generative AI comprehensively evaluates students' learning outcomes by synthesizing data from multiple sources, including assignment completion, classroom performance, and test scores. It can analyze students' mastery of individual knowledge points and skill areas in detail, such as vocabulary growth, improvement in grammatical application, and progress in oral expression. At the same time, it can provide specific remedial measures for weak areas, such as recommending relevant learning materials or developing targeted practice plans. This automated assessment and feedback mechanism not only saves teachers time and effort but also enables students to understand their learning status in a timely manner and adjust their learning strategies accordingly.

Generative AI can assess students' language proficiency across multiple dimensions. In addition to traditional measures of language knowledge and skills—such as reading, listening, speaking, and writing—it can evaluate students' language application ability, thinking skills, and cross-cultural communicative competence. By analyzing factors such as dialogue fluency, linguistic accuracy and appropriateness, and adaptability to different contexts in actual communicative scenarios, generative AI provides a comprehensive assessment of students' language application proficiency. Furthermore, through students' performance in group discussions or project-based tasks, it can assess their logical reasoning abilities and cross-cultural communicative competence. Generative AI enables a more comprehensive and in-depth understanding of students' learning status, identifies potential learning problems, and provides a strong basis for teachers to adjust teaching strategies and optimize teaching content, thereby better promoting students' holistic development and language proficiency improvement.

### 5. Challenges in AI-Empowered International Chinese Language Teaching

#### 5.1 Technical Adaptability of Teachers and Students

In international Chinese language teaching empowered by generative AI, the role of the teacher is shifting from knowledge transmitter to learning facilitator. Nevertheless, teachers must still maintain overall control of the classroom. Currently, content generated by generative AI may exhibit issues such as stereotypes and bias, overly templated responses that lack the flexibility to accommodate authentic, individualized needs, fabricated content, or privacy violations. This necessitates that teachers manually filter and decide which content is appropriate for classroom use. When using generative AI for learning, students need to have a clear understanding of learning directions and methods to ensure effective use.

### 5.2 Insufficient Affective Computing of Generative AI

Although generative AI is capable of providing immediate and rapid feedback, its feedback lacks the emotional expression of a real teacher. In the classroom, teachers can adjust their teaching strategies based on students' personalities and learning states. However, can AI recognize and adapt to learners' anxiety emotions? The lack of affective capacity in AI may negatively impact students' learning motivation. Particularly when correcting errors, the "mechanical feedback" of AI can easily trigger frustration and reduce learning drive.

### 5.3 Diminishment of Students' Creativity

Over-reliance on AI may weaken students' ability to think independently and their creativity. Take Chinese writing as an example. Chinese is a highly complex and unique language, rich in semantics, complex grammatical structures, and profound cultural connotations, with endless variations in linguistic expression. However, the model essays generated by generative AI tend to be standardized and templated. If students rely too heavily on AI, they will find it difficult to develop their own personal linguistic style and create original content.

## 6. Conclusion

The empowerment of generative AI has opened new horizons for international Chinese language teaching while also posing new challenges. Various AI models are powerful and have their respective strengths, yet they also face issues such as AI hallucinations and ethical challenges. In the future, there remains considerable room for development in areas such as dialect learning, the construction of customized intelligent agents, and the widespread availability of software and hardware. As technology advances rapidly, we must persist in learning and inquiry, and look forward to more constructive outcomes from scholars in the field of generative AI-empowered international Chinese language teaching.

## References

- Cai, W., Chu, C. Z., Cui, X. L., et al. (2025). DeepSeek-empowered innovation and development of international Chinese language education. *Chinese Teaching in the World*, 39(3), 291–315.
- Chen, L., & Zheng, S. (2025). Application of generative artificial intelligence in international Chinese writing instruction. *e-Education Research*, 46(9), 122–128.
- Du, M., & Zheng, D. X. (2025). Analysis of generative artificial intelligence reshaping digital and intelligent innovation in international Chinese language education. *Jilin University Journal Social Sciences Edition*, 65(1), 225–233, 240.
- Han, X. X., Ma, R. L., & Xu, J. (2025). Exploring the technical pathways of DeepSeek-empowered development of international Chinese teaching resources: A case study of graded reading text generation. *International Chinese Language Teaching Research*, (1), 30–40.
- Liu, Y. P., Ou, Z. G., & Wu, X. Q. (2025). Evaluation of the effectiveness of generative artificial intelligence in empowering international Chinese language teaching: A case study of instructional design, HSK mock test generation, and essay scoring. *Journal of Research on Education for Ethnic Minorities*, 36(1), 156–166.
- Ma, R. L., & Xu, J. (2025). DeepSeek-empowered innovations and transformations of the "five components" in smart international Chinese language education. *Journal of Yunnan Normal University (Teaching & Studying Chinese as a Foreign Language Edition)*, 23(2), 1–10.
- Ou, Z. G., Liu, Y. P., Qin, K., et al. (2024). Generation and evaluation of AI-based multimodal teaching resources: Application of AIGC in international Chinese language education. *Modern Educational Technology*, 34(9), 37–47.

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

# The Limited Classroom Application of Teacher Digital Literacy Training and Its Improvement Paths

Yuan Li<sup>1</sup>

<sup>1</sup> Hubei University of Education, Hubei, China

Correspondence: Yuan Li, Hubei University of Education, Hubei, China.

doi:10.63593/RAE.2788-7057.2026.06.003

## Abstract

Teacher digital literacy training is an important part of teacher professional development, but its effect is often limited when training content is not fully transferred into classroom practice. Teachers may learn how to use digital tools but still find it difficult to connect them with lesson goals, subject content, student needs, and real classroom conditions. This paper examines the main forms of limited classroom application and the reasons behind them. The discussion focuses on the separation between training and teaching needs, the strong focus on tool operation, limited practice opportunities, weak follow-up support, completion-based evaluation, and the practical costs of digital teaching. The paper argues that training should start from real classroom problems, connect digital skills with subject teaching, provide repeated practice, offer continued support, and evaluate changes in teaching practice rather than course completion alone. Teachers also need enough space to decide when and how digital tools should be used.

**Keywords:** teacher digital literacy, teacher training, classroom application, professional development, digital teaching

## 1. Introduction

Digital literacy is now an important part of teacher professional development. Although training can improve teachers' technical skills, many teachers still struggle to apply what they learn in daily classroom teaching. They may use digital tools only for basic tasks or return to familiar methods after a short trial.

This gap exists because classroom technology use depends on lesson goals, student needs, available resources, time, and classroom conditions. Training often focuses on how tools work rather than how they can solve specific teaching problems. Therefore, digital literacy training should be evaluated not only by participation or technical performance, but also by whether teachers can use technology appropriately and consistently in real classrooms. This paper examines the reasons for this transfer gap and discusses ways to improve training content, practice, follow-up support, and evaluation.

## 2. Teacher Digital Literacy Training and Classroom Application

### 2.1 The Main Goals of Teacher Digital Literacy Training

Teacher digital literacy training is not limited to teaching teachers how to operate devices or software. Its basic purpose is to help teachers use digital technology in ways that support teaching. This requires both technical skills and the ability to make suitable choices in real teaching situations.

Basic digital knowledge and skills remain an important part of training. Teachers need to know how to use common teaching platforms, digital devices, presentation tools, online resources, and communication systems. They also need basic skills in file management, information search, data use, and online communication. Without these skills, even simple digital teaching tasks can become difficult. However, technical ability is only

the starting point. Knowing how to use a tool does not mean knowing how to use it in a lesson.

The selection of digital teaching resources is another important part of digital literacy. Online resources are easy to access, but their quality varies greatly. Teachers need to judge whether a resource is accurate, suitable for the lesson, and appropriate for the age and learning level of students. A video, online exercise, or digital textbook may contain useful information, but it may not match the purpose of the lesson. Teachers also need to avoid adding digital content only because it is attractive or convenient. Resource selection should be linked to what students need to learn.

Digital lesson design requires a stronger connection between technology and teaching. Teachers need to decide where digital tools should appear in a lesson and what role they should play. A tool may be used to explain a difficult concept, organize classroom interaction, provide practice, collect student responses, or support review. These uses require different forms of lesson design. If technology is added after the lesson structure has already been decided, it can easily become an extra part with little teaching value. Training should help teachers think about technology while they are planning the lesson, rather than treating it as a separate technical step.

Teacher digital literacy also includes the use of technology to support student learning. Digital tools can provide different types of materials, allow students to work at different speeds, and give teachers access to learning records and student responses. These functions may help teachers identify common mistakes or provide more focused support. Digital assessment is also part of this process. Online quizzes, learning platforms, and digital records can provide useful information, but teachers still need to interpret the results. A score or data report cannot explain every learning problem. Teachers must connect the information with classroom observation and their knowledge of students.

Responsible use of digital technology is equally important. Teachers need to consider student privacy, information security, copyright, screen time, and the reliability of online content. They also need to understand that not every digital tool is suitable for every teaching task. Some lessons may work better through direct explanation, discussion, reading, writing, or hands-on activities. Digital literacy includes the ability to decide when technology is useful and when it is not necessary.

## *2.2 What Classroom Application Means*

Classroom application should not be measured by how many digital tools a teacher uses. A classroom filled with devices, platforms, and digital materials is not necessarily a digitally effective classroom. The more important issue is whether technology helps the teacher solve a teaching problem or helps students understand, practice, discuss, or apply what they are learning.

This distinction matters because digital teaching can remain at a very simple level. A teacher may replace a blackboard with presentation slides or play a video instead of giving an oral explanation. These changes make technology visible, but the basic teaching process may remain the same. In such cases, digital tools change the form of teaching more than the learning process. The classroom use of technology becomes more meaningful when it changes how teachers organize learning or respond to student needs.

Classroom application also depends on teaching objectives. Teachers need to know what students should learn before deciding which technology to use. If the goal is to help students understand an abstract idea, a visual simulation may be useful. If the goal is to improve discussion and argument, a digital response tool may help collect different views. If students need repeated practice, an online exercise system may provide quick feedback. The value of the tool comes from its connection with the task, not from the tool itself.

Subject content also affects the way digital technology is used. The same tool may have different value in different subjects. A digital map may be useful in geography, while a simulation may be more suitable for science. Language teaching may use audio, video, or collaborative writing tools. Mathematics teaching may use visual models or immediate practice feedback. General training can introduce common tools, but teachers still need to adapt them to the logic and needs of their own subjects.

Student differences are another part of classroom application. Students do not have the same level of digital skill, learning ability, attention, or access to devices. A digital activity that works well for one class may not work in another. Teachers need to judge whether students can use the tool without being distracted and whether the activity helps those who learn at different speeds. They may also need to prepare non-digital alternatives when equipment or access is limited.

Classroom conditions can further shape teachers' choices. Time, internet quality, available devices, class size, and school rules all affect whether a digital method can be used smoothly. A method demonstrated in training may require stable equipment and enough preparation time. In daily teaching, these conditions may not always be available. Teachers often have to simplify or adjust what they learned.

For this reason, classroom application should be understood as a change in teaching decisions and teaching

practice. A teacher who has completed digital literacy training should be able to judge whether a tool fits the lesson, adapt it to the classroom, and stop using it when it does not improve learning. The key point is not frequent technology use, but appropriate use based on teaching goals, subject needs, student conditions, and classroom reality.

### **3. Main Problems in the Classroom Application of Digital Literacy Training**

#### *3.1 Training Content Is Separated from Real Classroom Needs*

A common problem in digital literacy training is the gap between what teachers learn and what they need in daily teaching. Many training courses spend considerable time introducing software functions, online platforms, digital resources, and operating procedures. Teachers learn where to click, how to upload materials, how to create online activities, or how to use a new application. These skills are useful, but they do not directly answer the questions teachers face in class.

Classroom problems are usually more specific. A teacher may need to explain a difficult concept, keep students focused, support students with different learning levels, or check whether students have understood the lesson. In these situations, knowing how a digital tool works is not enough. Teachers also need to know whether the tool fits the problem and how it should be used without interrupting the lesson.

This gap becomes clear when teachers return to their own classrooms. They may remember the steps taught in training but still be unsure where the tool fits into a lesson. Some tools may require more preparation time than teachers can provide. Others may not suit the subject, the age of the students, or the classroom environment. Teachers then use only the simplest functions or stop using the tool after a short period.

The problem is not always a lack of digital ability. Teachers may already have enough technical knowledge. What is missing is a clear link between that knowledge and daily teaching tasks. When training gives limited attention to real classroom problems, teachers have to make this connection by themselves.

#### *3.2 Training Emphasizes Learning Tools More Than Designing Teaching*

Digital literacy training often treats technology as a separate skill. A course may be organized around a particular platform, application, or type of software. Teachers are shown its main functions and asked to complete several practice tasks. This makes the training easy to organize, but it can place too much attention on the tool itself.

Teaching does not begin with a digital tool. It begins with decisions about what students need to learn, what difficulties they may have, and what teaching activities can help them. Technology should be selected after these questions are considered. When training starts from the tool, teachers may try to find a place for it in the lesson even when it is not needed.

The weak connection between digital skills and lesson design can also lead to separate digital activities. A teacher may add an online quiz, a short video, or an interactive task because these methods were introduced during training. The activity may look different from traditional teaching, but it may have little connection with the main learning goal. Students complete the activity, yet the teacher may not use the results to adjust later teaching.

Subject differences receive limited attention in some general training programs. The same digital tool is introduced to teachers from different subject areas, even though their teaching tasks are not the same. A tool that supports visual explanation may be useful in one subject but have little value in another. Without subject-based examples and lesson design practice, teachers may find it difficult to judge how a general digital skill can be used in their own teaching.

#### *3.3 Training Tasks Differ from Everyday Teaching Conditions*

Training activities are often completed under conditions that are easier to control than normal classroom teaching. The equipment is prepared in advance, internet access is stable, the task has a clear structure, and participants know what they are expected to do. Trainers can also provide immediate help when technical problems appear. These conditions make it easier to demonstrate the value of a digital tool.

Daily classrooms are less predictable. Teachers have limited lesson time and must deal with student questions, different learning speeds, classroom order, equipment problems, and changes in the teaching schedule. A digital activity that works smoothly during training may take too much time in a real class. Students may have trouble logging in, devices may not work as expected, or an online task may take longer than planned.

Teachers also have to consider the cost of preparation. A digital lesson may require teachers to search for materials, prepare online tasks, test links, check devices, and prepare a backup plan. These tasks add to normal lesson preparation. If the teaching benefit is small or uncertain, teachers may decide that the extra work is not worthwhile.

Demonstration tasks can also create expectations that are difficult to meet. Training examples are often carefully

prepared and show a smooth teaching process. Real lessons rarely follow such a clear pattern. Student responses may differ from what the teacher expected, and the teacher may need to change the activity during the lesson. When teachers are trained mainly through ideal examples, they receive little guidance on how to deal with these common classroom problems.

### *3.4 Classroom Use Often Remains at a Surface Level*

Limited classroom transfer does not always mean that teachers stop using digital tools. In many cases, they continue to use technology, but only for simple tasks. Presentation slides, videos, digital textbooks, online resources, and basic response tools are common examples. These uses can make teaching more convenient, but they do not always change how students learn.

Presentation is a clear example. Teachers may replace writing on the board with slides or replace printed pictures with digital images. The form of delivery changes, but students may still mainly listen, watch, and record information. The teacher remains the main source of content, and students have little change in their role during the lesson.

A similar problem can occur with digital interaction. Teachers may use online quizzes, polls, or response systems to make the classroom appear more interactive. If the answers are collected but not used to identify learning problems or adjust teaching, the activity adds a digital step without changing the learning process. Technology is present, but its teaching value remains limited.

Surface-level use can also result from the way digital teaching is understood. When visible technology use is treated as a sign of good digital teaching, teachers may feel pressure to include digital tools even when the lesson does not require them. Attention then shifts from whether students learn better to whether technology appears in the classroom.

Some teachers also stay with simple uses because they are easier to control. Playing a video or showing slides carries less risk than reorganizing a lesson around student collaboration, digital feedback, or different learning paths. Deeper use requires more planning and stronger classroom control. Without enough practice and support, teachers may choose forms of digital teaching that are familiar, stable, and easy to manage.

## **4. Why Classroom Transfer Remains Limited**

### *4.1 Training Design Gives Too Little Attention to Classroom Context*

Digital literacy training is often designed for a broad group of teachers. The same course may be used for teachers from different subjects, grade levels, and teaching environments. This makes training easier to organize, but it also reduces its relevance to daily classroom work. A language teacher, a science teacher, and a mathematics teacher may all learn the same digital tool, even though they use technology for very different teaching purposes.

Teachers also have few chances to bring their own classroom problems into the training process. Many courses use fixed examples prepared by trainers. These examples may explain the functions of a tool clearly, but they do not always match the problems teachers face in their own lessons. A teacher may want to know how to support students with different learning levels, how to use digital feedback in a large class, or how to avoid losing lesson time when using a platform. If these issues are not included in training, the teacher must find solutions later without much guidance.

Classroom context also affects whether a digital method can be used at all. Available devices, internet quality, class size, student age, lesson length, and subject content can all change the result. A method that works well in a small training group may be difficult to use in a crowded classroom. When these conditions are not considered during training, teachers may understand the method but still find it hard to apply.

### *4.2 Short-Term Training Provides Limited Opportunities for Practice*

Many digital literacy programs are short courses, workshops, or online learning tasks. They can introduce a new tool or method within a limited period, but classroom use requires more than understanding the basic steps. Teachers usually need to try the method several times before they know how to fit it into a lesson.

Some problems only appear during real teaching. An activity may take longer than expected. Students may misunderstand the instructions. A platform may work slowly. A digital task may create more classroom management problems than the teacher expected. These situations cannot be fully learned through demonstrations or short practice tasks.

Teachers also need time to adjust their lesson plans after each attempt. A method that does not work well the first time may become useful after changes in task design, timing, or student grouping. Short-term training rarely leaves enough space for this process. The training ends before teachers have tested the method in different lessons.

Reflection is also limited when training is compressed into a short period. Teachers may finish a task and move directly to the next topic. They have little time to think about what worked, what failed, and why. Without repeated use and reflection, new digital methods often remain temporary skills rather than part of regular teaching.

#### *4.3 Follow-Up Support After Training Is Weak*

Support often decreases once formal training ends. During the course, teachers can ask trainers for help and receive clear instructions. After returning to their classrooms, they may have to solve new problems alone.

Technical problems are only one part of this difficulty. Teachers may know how to fix a login problem or change a platform setting, but they may not know how to adjust the teaching activity itself. For example, students may focus too much on the device, the task may not produce useful responses, or the digital activity may take time away from more important parts of the lesson. These are teaching problems rather than simple technical problems.

Without follow-up support, small difficulties can reduce teachers' willingness to continue trying. If a method fails several times, teachers may decide that it is not suitable for their classes. They may return to familiar methods because these methods are easier to manage and involve less risk.

Peer support is also not always built into training programs. Teachers often face similar problems, but they may have few regular opportunities to discuss classroom use, compare lesson designs, or share adjustments that worked. In this situation, the experience gained by one teacher may remain personal rather than becoming useful to others.

#### *4.4 Evaluation Focuses More on Completion Than on Teaching Change*

The results of digital literacy training are often measured through indicators that are easy to record. These may include attendance, course hours, online learning records, test scores, certificates, or completed assignments. Such information can show whether teachers finished the training, but it says little about what happens in their classrooms later.

A teacher can complete a digital course successfully without changing daily teaching practice. Online assignments may also be completed mainly to meet course requirements. If the evaluation ends when the training ends, there is little reason to examine whether the new knowledge is used after the course.

This form of evaluation can also affect how teachers approach training. When course completion becomes the main goal, teachers may pay more attention to meeting formal requirements than to testing new methods in class. They learn what is needed for the assessment, but classroom use may remain limited.

Teaching change is harder to measure because it takes time. It may involve changes in lesson design, classroom interaction, feedback, student participation, or the teacher's choice of resources. These changes are not easily shown by a certificate or a test score. They require classroom evidence and continued observation after training.

#### *4.5 Teachers Face Practical Costs When Trying New Digital Practices*

Using a new digital method usually increases the amount of preparation needed at the beginning. Teachers may need to search for suitable materials, create digital tasks, test links, check devices, prepare student instructions, and think about possible technical problems. They may also need a backup plan in case the system does not work.

These tasks take time. Teachers already have lesson preparation, marking, student support, meetings, and other duties. If a digital method requires much more preparation but brings only a small improvement in learning, it is difficult to use it regularly.

Technical uncertainty also creates pressure. Even a well-prepared digital lesson may be interrupted by unstable internet access, device failure, software updates, or student login problems. The teacher must then solve the problem while keeping the class under control. This risk can make familiar teaching methods more attractive.

Teachers also consider whether a new method is worth the effort. If they cannot clearly see how it improves student understanding, participation, or feedback, they may not continue using it. A digital activity may look new, but if it adds extra work without solving a real teaching problem, it is unlikely to become part of everyday practice.

### **5. Improvement Paths for Stronger Classroom Application**

#### *5.1 Start Training from Real Classroom Problems*

Digital literacy training should begin with teaching problems rather than digital tools. Teachers usually care less about what a platform can do than about whether it can help them solve a problem in class. Training topics can therefore be built around lesson planning, classroom interaction, student assessment, homework feedback, and

support for students with different learning needs.

This approach changes the role of technology in training. A tool is no longer introduced as the main content of the course. It becomes one possible way to deal with a teaching task. For example, a training activity on classroom feedback can begin with the problem of finding out which students have understood a concept. Teachers can then compare different methods, including digital response tools, short quizzes, or non-digital questioning. The focus remains on the teaching purpose.

Teachers should also be encouraged to bring their own classroom cases into training. A teacher may have difficulty keeping students involved in group work or giving timely feedback in a large class. These problems can become training tasks. Teachers can design a possible solution, try it, and discuss whether the digital method is useful. This makes the training more closely connected with daily teaching.

### *5.2 Connect Digital Skills with Subject Teaching*

General digital skills need to be linked with subject teaching. Teachers from different subjects do not use technology in the same way. A tool that works well in science may have limited value in language teaching. A useful training program should show how digital methods can support the learning goals of different subjects.

Subject-based examples can make this connection clearer. Language teachers may need digital tools for listening practice, writing feedback, or text discussion. Mathematics teachers may use visual models, graphing tools, or practice systems. Science teachers may use simulations, videos, or digital experiments. Art and music teachers may need tools for creation, recording, or performance review.

Training tasks should also require teachers to explain why a tool is suitable for a particular lesson. This helps move attention away from tool functions. Teachers need to think about the learning goal, the type of student activity, and the value of the digital method. If a simpler method works better, there is no need to use technology.

Subject teaching can also be used as the main setting for training practice. Teachers can work on a real unit or lesson from their own subject rather than completing a general digital task. This gives them a clearer idea of how the method can be used after the training ends.

### *5.3 Add Classroom Practice to the Training Process*

Teachers need chances to use new digital methods in real classrooms during the training period. Demonstrations and practice tasks can show how a tool works, but they cannot fully prepare teachers for the problems that appear in daily teaching.

Classroom practice does not need to begin with a large change. Teachers can test one small method in one part of a lesson. They may use a digital quiz to check understanding, try an online discussion task, or use a shared document for group work. A small trial creates less pressure and makes it easier to see what needs to be changed.

The experience from the trial should return to the training process. Teachers can discuss what happened in class, what problems appeared, and how students responded. They can then revise the activity and try it again. This cycle of design, trial, discussion, and revision gives teachers time to turn a new method into a usable classroom practice.

Repeated use is important because the first attempt may not work well. A teacher may choose the wrong task length, give unclear instructions, or use a tool that does not fit the students. These problems are easier to correct when the training allows more than one classroom trial.

### *5.4 Build Continuous Support After Formal Training*

Support should not stop when a workshop or course ends. Many teachers meet the most difficult problems only after they begin using a new method on their own. They may need help with lesson design, classroom management, student response, or technical issues.

Schools can provide regular spaces for teachers to discuss these problems. Small teaching groups can share examples, compare lesson plans, and talk about failed attempts as well as successful ones. This type of discussion is often more useful than another general introduction to a new tool.

Mentoring can also help teachers who are less confident with digital teaching. A more experienced teacher can observe a lesson, review a task, or suggest a simpler way to use a tool. The aim is not to create a fixed model for everyone. The support should help teachers adjust a method to their own classroom.

Shared resources can reduce repeated work. Teachers often spend time searching for materials or creating similar digital tasks. A common resource space can include lesson examples, activity templates, technical guides, and notes about common problems. These materials are more useful when they come from real classroom use rather than from tool descriptions alone.

### *5.5 Change the Evaluation of Training Effectiveness*

Training evaluation should pay more attention to classroom use. Attendance, course hours, test scores, and certificates can show whether teachers completed the program, but they cannot show whether teaching practice changed.

Lesson plans can provide some evidence of transfer. They can show whether teachers are able to connect digital tools with learning goals and student activities. Classroom observation can provide more direct information. It can show whether technology supports teaching or simply adds another step to the lesson.

Teacher reflection is also useful. After using a digital method, teachers can record what worked, what caused problems, and what they would change next time. These reflections can help trainers understand whether the course content is suitable for real classrooms.

Student activity should also be considered. The value of digital teaching is not shown by the number of tools used. It is more important to see whether students receive clearer feedback, take part in useful interaction, understand difficult content more easily, or gain more chances to practice. A teacher who uses one digital tool effectively may show stronger classroom transfer than a teacher who uses many tools without a clear purpose.

#### *5.6 Give Teachers Space to Adapt Digital Practices*

Digital teaching should not become a fixed model that every teacher is expected to copy. Teachers work with different subjects, students, class sizes, schedules, and resources. The same method may produce different results in different classrooms.

Teachers need room to change what they learned during training. They may shorten an activity, replace one tool with another, combine digital and non-digital methods, or decide not to use technology in a particular lesson. These choices are part of professional teaching rather than signs that the training has failed.

Training should also help teachers judge the limits of digital methods. Some tasks can be completed more clearly through direct explanation, discussion, reading, writing, or hands-on work. A digital tool should be used when it adds value to the lesson, not because its use is expected.

Giving teachers more choice can also make digital practices easier to maintain. When teachers are allowed to select methods that fit their own teaching style and classroom conditions, they are more likely to continue using them. The purpose of digital literacy training is to improve teachers' ability to make suitable decisions about technology, not to make every classroom look the same.

### **6. The Boundary of Effective Digital Literacy Training**

Digital literacy training should not encourage teachers to use more technology simply for the sake of using it. A lesson is not better because it includes more digital tools. The real question is whether the technology helps students understand the content, take part in the lesson, receive useful feedback, or complete a learning task more effectively.

Some teaching activities do not need digital support. Direct explanation may be more suitable when a teacher needs to clarify a difficult point. Reading, writing, discussion, experiments, and face-to-face interaction may also work better without extra digital steps. If technology makes the task more complicated or takes attention away from the learning goal, its use may have little value.

Teachers therefore need the ability to judge when a digital tool is useful. This judgment depends on the lesson goal, subject content, student needs, available time, and classroom conditions. A tool may be suitable in one lesson but unnecessary in another. Training should help teachers compare different options instead of giving the impression that digital methods are always preferred.

The value of training should also be judged by the quality of teachers' decisions. A teacher who chooses not to use a digital tool after considering the needs of the lesson may show stronger digital literacy than a teacher who uses several tools without a clear purpose. Effective digital teaching includes knowing the limits of technology.

For this reason, training needs to move beyond teaching teachers how to operate devices, platforms, and software. Teachers need to understand why a tool may be useful, when it should be used, how it can be adjusted, and when it should be replaced by another method. This type of judgment is more closely connected with classroom teaching than simple technical skill.

### **References**

- Avalos, B. (2011). Teacher professional development in *Teaching and Teacher Education* over ten years. *Teaching and Teacher Education*, 27(1), 10–20. <https://doi.org/10.1016/j.tate.2010.08.007>
- Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better

- conceptualizations and measures. *Educational Researcher*, 38(3), 181–199. <https://doi.org/10.3102/0013189X08331140>
- Ertmer, P. A., Ottenbreit-Leftwich, A. T., Sadik, O., Sendurur, E., & Sendurur, P. (2012). Teacher beliefs and technology integration practices: A critical relationship. *Computers & Education*, 59(2), 423–435. <https://doi.org/10.1016/j.compedu.2012.02.001>
- Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice*, 8(3), 381–391. <https://doi.org/10.1080/135406002100000512>
- Ministry of Education of the People's Republic of China. (2022). 教师数字素养 [Teacher digital literacy] (JY/T 0646–2022).
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Philipsen, B., Tondeur, J., Pareja Roblin, N., Vanslambrouck, S., & Zhu, C. (2019). Improving teacher professional development for online and blended learning: A systematic meta-aggregative review. *Educational Technology Research and Development*, 67(5), 1145–1174. <https://doi.org/10.1007/s11423-019-09645-8>
- Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu* (Y. Punie, Ed.). Publications Office of the European Union. <https://doi.org/10.2760/159770>
- Tondeur, J., van Braak, J., Sang, G., Voogt, J., Fisser, P., & Ottenbreit-Leftwich, A. (2012). Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence. *Computers & Education*, 59(1), 134–144. <https://doi.org/10.1016/j.compedu.2011.10.009>
- UNESCO. (2018). *UNESCO ICT competency framework for teachers* (Version 3). UNESCO.

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

# Are We Really Listening? Development and Psychometric Validation of a Five-Stage Listening Process Scale for Higher Education Students in Region II, Philippines

Mark Kevin Astrero<sup>1</sup>, Saranay I. Doyaoen<sup>1</sup>, Darius King Galo<sup>1</sup>, Aileen Melody De Vera<sup>1</sup> & Titov Dela Cruz<sup>2</sup>

<sup>1</sup> Northeastern College, Santiago City, Isabela, Philippines

<sup>2</sup> University of La Salette Incorporated High School, Santiago City, Isabela, Philippines

Correspondence: Mark Kevin Astrero, Northeastern College, Santiago City, Isabela, Philippines.

doi:10.63593/RAE.2788-7057.2026.06.004

## Abstract

Listening is central to classroom learning yet rarely measured with locally validated tools in Philippine higher education. This study developed and validated a 50-item Listening Process Scale (LPS), grounded in Brownell's stage model (receiving, understanding, remembering, evaluating, responding), using 314 students from higher education institutions in Region II, Philippines. Sampling adequacy was excellent (KMO = .968). Exploratory and confirmatory factor analyses converged on a hierarchical structure: five correlated first-order factors loading strongly (.88–.91) onto one General Listening Process factor, rather than five independent stages. Reliability was excellent overall ( $\alpha = .976$ ) and per subscale (.900–.937), though elevated inter-item correlations suggest some redundancy. Structural equation modeling showed the General factor significantly predicted academic performance ( $\beta = .212$ ,  $p < .001$ ). Students self-reported High listening levels across all stages ( $M = 3.59$ ). The LPS is a reliable, academically relevant measure, best interpreted through its Overall score pending further item refinement.

**Keywords:** listening process, questionnaire validation, exploratory factor analysis, HURIER model, higher education institutions, Region II Philippines, scale development

## 1. Introduction

Among the four macro-communication skills — listening, speaking, reading, and writing — listening is simultaneously the most frequently used and the least deliberately taught (Wolvin & Coakley, 1996; Gilakjani & Ahmadi, 2011). College and university students spend a substantial share of their academic day receiving spoken information through lectures, discussions, and peer interaction, yet listening competence is seldom assessed with the same rigor applied to reading comprehension or written expression. This asymmetry is consequential: instructors, program chairs, and student-support units in Philippine higher education institutions (HEIs) have few locally validated instruments with which to diagnose where students' listening difficulties actually lie — at the point of attention, comprehension, retention, evaluation, or response.

The present study addresses this gap by developing and validating a Listening Process Scale (LPS) for HEI students in Region II (Cagayan Valley), Philippines. Rather than treating listening as a single undifferentiated trait, the LPS operationalizes listening as a sequence of five interrelated cognitive-behavioral stages — receiving/attending, understanding, remembering, evaluating, and responding — consistent with stage models of the listening process that have long informed communication pedagogy (Brownell, 1985, 1996). Validating such a stage-based instrument locally serves two purposes. First, it gives Philippine HEIs a diagnostic tool sensitive enough to identify which stage of the listening process students struggle with most, rather than a single global score that obscures where intervention should be targeted. Second, it contributes empirical evidence, from an

under-represented Southeast Asian population, to the broader question of whether listening stage models replicate as separable statistical dimensions or collapse into fewer, broader factors when measured with self-report items — a question with direct relevance to how listening curricula and assessments should be structured.

Specifically, this study sought to: (1) examine the factorial structure of the 50-item LPS using exploratory factor analysis; (2) evaluate the internal consistency reliability of the overall scale and its five hypothesized subscales; (3) describe the self-reported level of each listening stage among Region II HEI students; and (4) examine whether LPS scores differ meaningfully across student sex, age bracket, first-language background, and academic performance, as a source of criterion-related validity evidence.

## **2. Theoretical Framework and Related Literature**

### *2.1 Stage Models of the Listening Process*

Brownell's HURIER model remains one of the most widely cited frameworks for conceptualizing listening as a multi-stage process rather than a single skill, decomposing listening into hearing, understanding, remembering, interpreting, evaluating, and responding (Brownell, 1985, 1996). The model treats these stages as interdependent rather than strictly sequential: weaknesses at an early stage (e.g., attentional lapses during hearing) constrain what can be accomplished at later stages (e.g., accurate evaluation or an appropriate response). The present instrument adapts this tradition into five measurable dimensions — collapsing hearing/attending into a single receiving stage and folding interpretive judgments (attending to tone, context, and underlying intent) into the understanding stage — a simplification intended to make the construct tractable for self-report measurement with a student population while preserving the process logic of the original model.

### *2.2 Measuring Listening: Prior Instruments and Open Questions*

Several validated listening instruments exist in the communication literature, including the Watson–Barker Listening Test (Bodie, Worthington, & Fitch-Hauser, 2011), the Listening Styles Profile–16 and its revision (Bodie & Worthington, 2010; Bodie, Worthington, & Gearhart, 2013), and the Active-Empathic Listening Scale (Bodie, 2011). Collectively, this body of work has repeatedly found that self-report listening measures are prone to producing a large, dominant general factor, with narrower conceptually distinct facets emerging only when items are written with sufficient specificity and sample sizes are adequate (Bodie, Worthington, Imhof, & Cooper, 2008; Bodie, 2013). This measurement literature — developed almost entirely on North American samples — has not, to the researchers' knowledge, been extended to a stage-based instrument validated on a Filipino HEI population, which the present study addresses.

### *2.3 Listening and Academic Performance in the Philippine Context*

Locally, listening difficulties have been repeatedly linked to broader academic and language-learning outcomes. Descriptive studies of Philippine secondary and tertiary learners report that listening proficiency lags behind other language skills and correlates with academic indicators such as general weighted average (De Vera et al., 2018, as cited in academic listening research on Philippine learners; Aguacito, Macatabon, & Imbang-Agni, 2025). These studies, however, have generally relied on researcher-made instruments without a formal psychometric validation stage, which limits the confidence with which their findings can be generalized or their instruments reused. The present study responds to this methodological gap by subjecting a stage-based listening instrument to a full sequence of sampling-adequacy testing, exploratory factor analysis, reliability analysis, and known-groups validation.

## **3. Methods**

### *3.1 Research Design*

This study employed a descriptive-validation, cross-sectional survey design, combining psychometric scale-validation procedures with a descriptive assessment of self-reported listening behavior.

### *3.2 Participants and Sampling*

Participants were 314 students drawn from higher education institutions across Region II, Philippines, recruited through an online, self-administered questionnaire distributed via institutional and social-media channels (convenience sampling). Table 1 summarizes the demographic profile of the sample. The sample was predominantly female (62.4%) and in the 19–21 age bracket (58.0%); most respondents identified Filipino/Tagalog as their family's first language (67.5%), and the modal academic-performance bracket was 85–89 (43.6%).

Table 1. Demographic Profile of Respondents

| Profile Variable                       | Category             | n   | %    |
|----------------------------------------|----------------------|-----|------|
| Sex                                    | Female               | 196 | 62.4 |
|                                        | Male                 | 118 | 37.6 |
| Age Bracket                            | 16–18                | 103 | 32.8 |
|                                        | 19–21                | 182 | 58.0 |
|                                        | 22 and above         | 29  | 9.2  |
| Family's First Language                | Filipino/Tagalog     | 212 | 67.5 |
|                                        | Iloco                | 30  | 9.6  |
|                                        | Visaya               | 29  | 9.2  |
|                                        | English              | 26  | 8.3  |
|                                        | Other local dialects | 17  | 5.4  |
| Academic Performance (General Average) | below 80             | 11  | 3.5  |
|                                        | 80–85                | 71  | 22.6 |
|                                        | 85–89                | 137 | 43.6 |
|                                        | 90–100               | 95  | 30.3 |

### 3.3 Instrument

The LPS is a 50-item, five-point Likert scale (1 = strongly disagree/almost never, 5 = strongly agree/almost always) composed of five researcher-defined, 10-item subscales corresponding to the receiving/attending, understanding, remembering, evaluating, and responding stages of listening. Items were written in behaviorally specific, first-person statements (e.g., “I give my full attention to the speaker without being distracted by my phone or other devices”; “I recall the key points of a conversation shortly after it ends”) to reduce the abstraction typical of trait-style listening items.

### 3.4 Data Quality Screening

Because the questionnaire was self-administered online without proctoring, response patterns were screened for evidence of careless or inattentive responding prior to the factor-analytic and reliability analyses, following recommended practice for online survey data (Meade & Craig, 2012). Twenty-three respondents (7.3% of the sample) provided the identical rating to all 50 items — a classic “straight-lining” pattern that artificially inflates inter-item correlations and can distort factor solutions. These 23 cases were retained for the descriptive-level and reliability analyses (consistent with common practice, since Cronbach’s alpha and mean-level description are comparatively robust to this pattern) but were excluded from the exploratory factor analysis and the Kaiser–Meyer–Olkin/Bartlett sampling-adequacy tests, which are more sensitive to spurious covariance; the resulting factor-analytic sample was  $N = 291$ . Both the full-sample and screened-sample results are reported for transparency.

### 3.5 Data Analysis

Sampling adequacy was evaluated using the Kaiser–Meyer–Olkin (KMO) measure and Bartlett’s test of sphericity. Because parallel analysis has been shown to outperform simpler heuristics such as the Kaiser eigenvalue-greater-than-one rule for determining the number of factors to retain (Watkins, 2018), the number of factors was determined using Horn’s parallel analysis (100 random-data replications) before extracting factors via exploratory factor analysis (EFA) using minimum-residual (minres) estimation with promax (oblique) rotation, appropriate given the theoretically correlated nature of sequential listening stages. Mean inter-item and inter-scale correlations were computed to characterize item redundancy and discriminant separation among subscales (Clark & Watson, 1995).

The measurement structure suggested by the EFA was then formally tested with confirmatory factor analysis (CFA), estimated by maximum likelihood in the Python package *semopy* (Igolkina & Meshcheryakov, 2020). Four nested models were compared: (1) the a priori five-factor correlated CFA model corresponding to the original subscale design; (2) an EFA-informed five-factor correlated CFA model built from the empirical item groupings; (3) a second-order (hierarchical) CFA model with the five first-order factors loading onto a single General Listening Process factor; and (4) a full structural equation model (SEM) extending the second-order model with a structural regression path from the General factor to students’ academic performance,

operationalized as the numeric midpoint of their reported general-average bracket. Model fit was evaluated using the chi-square test, Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA), following conventional joint-criteria guidelines (CFI/TLI  $\geq .90$  acceptable,  $\geq .95$  good; RMSEA  $\leq .08$  acceptable,  $\leq .06$  good; Hu & Bentler, 1999).

Internal consistency was assessed with Cronbach’s alpha for the overall scale and each subscale, interpreted with attention to the caution that very high alpha values can themselves signal item redundancy rather than pure reliability (Taber, 2018). Because Shapiro–Wilk testing indicated departure from normality in the composite scores, group comparisons used the nonparametric Mann–Whitney U test (two-group comparisons: sex) and the Kruskal–Wallis H test (comparisons across three or more groups: age bracket, first language, academic performance). All analyses were conducted in Python 3.12 using pandas, SciPy, factor\_analyzer, and semopy.

### 3.6 Ethical Considerations

Participation was voluntary and anonymous; respondents’ email addresses, collected only for survey-platform administration, were removed prior to analysis and are not reported. No personally identifying information is presented in this manuscript.

## 4. Results

### 4.1 Sampling Adequacy

The overall Kaiser–Meyer–Olkin measure of sampling adequacy was .968 for the full sample ( $N = 314$ ) and .963 for the screened sample ( $N = 291$ ), both well above the conventional “marvelous” threshold of .90 (Watkins, 2018). Bartlett’s test of sphericity was significant for both samples (full sample:  $\chi^2(1225) = 11,411.63$ ,  $p < .001$ ; screened sample:  $\chi^2(1225) = 9,718.90$ ,  $p < .001$ ), confirming that the item correlation matrix was suitable for factor analysis.

### 4.2 Number of Factors and Factor Structure

Horn’s parallel analysis on the screened sample indicated that only the first three empirical eigenvalues exceeded the 95th percentile of eigenvalues generated from random data of the same dimensions (real eigenvalues of 21.75, 2.47, and 1.95 versus random-data thresholds of 1.98, 1.87, and 1.80, respectively), suggesting a strong general listening factor with two additional, narrower factors. The conventional Kaiser criterion (eigenvalue  $> 1$ ) was considerably more liberal, suggesting seven factors — illustrating the well-documented tendency of that criterion to over-extract (Watkins, 2018).

Given the strength of the theoretical five-stage model and the interpretability criterion recommended alongside statistical stopping rules (Watkins, 2018), a five-factor solution was also examined and is reported as the primary structural result, with the three-factor parallel-analysis solution noted as a more conservative alternative. The five-factor promax solution explained 43.2% of cumulative variance (Factor 1 = 12.6%, Factor 2 = 11.1%, Factor 3 = 8.2%, Factor 4 = 7.3%, Factor 5 = 4.0%) and produced the pattern summarized in Table 2.

Table 2. Summary of the Five-Factor Promax Solution

| Empirical Factor                    | Items (primary loading $\geq .29$ ) | n items | Dominant a priori stage(s)                |
|-------------------------------------|-------------------------------------|---------|-------------------------------------------|
| Factor 1 — Focused Reception        | Q2–Q20 (19 items)                   | 19      | Receiving/Attending + Understanding       |
| Factor 2 — Critical Evaluation      | Q31–Q40, Q49                        | 11      | Evaluating                                |
| Factor 3 — Retention/Memory         | Q21, Q23–Q25, Q27–Q30               | 8       | Remembering                               |
| Factor 4 — Responsive Engagement I  | Q1, Q22, Q26, Q41, Q44–Q47, Q50     | 9       | Responding (+2 Remembering, +1 Receiving) |
| Factor 5 — Responsive Engagement II | Q42, Q43, Q48                       | 3       | Responding                                |

The pattern in Table 2 shows partial, rather than complete, correspondence with the a priori five-stage design. The receiving/attending and understanding stages, hypothesized as separate dimensions, loaded onto a single broad Focused Reception factor, suggesting that Region II HEI respondents did not distinguish, in their self-reports, between attending to a speaker and comprehending the speaker’s message — a plausible outcome given that both stages occur in rapid succession during real-time listening and may be difficult for respondents to introspectively separate. In contrast, the evaluating and remembering stages emerged as clean, largely unidimensional factors closely matching their hypothesized item sets. The responding stage split across two related factors (Factor 4 and Factor 5), both dominated by responding items but differing in emphasis — Factor

4 loading items describing embodied/nonverbal engagement and initiating verbal responses, and Factor 5 loading items describing follow-up questioning and constructive expression of disagreement. Six items (Q2, Q12, Q13, Q18, Q29, Q30) showed meaningful cross-loadings ( $> .32$ ) on more than one factor and are flagged as candidates for revision or removal in future refinements of the scale.

#### 4.3 Inter-Item and Inter-Scale Correlations

Table 3 summarizes mean inter-item correlations within each subscale, computed on the full sample ( $N = 314$ ). Mean within-subscale inter-item correlations ranged from .476 (Receiving/Attending) to .598 (Evaluating), with an overall mean inter-item correlation of .453 across all 50 items. These values sit above the .15–.50 range conventionally recommended for unidimensional item sets (Clark & Watson, 1995) — a pattern consistent with the very high alpha coefficients reported in Section 4.6 and with the dominant general factor identified in the parallel analysis, and reinforcing the interpretation that several items within each subscale are highly redundant rather than each capturing a distinct facet of that stage.

Table 3. Mean Inter-Item Correlations Within Subscales

| Subscale            | k (items) | Mean r | Min r | Max r |
|---------------------|-----------|--------|-------|-------|
| Receiving/Attending | 10        | .476   | .355  | .572  |
| Understanding       | 10        | .522   | .302  | .689  |
| Remembering         | 10        | .515   | .193  | .644  |
| Evaluating          | 10        | .598   | .487  | .715  |
| Responding          | 10        | .547   | .326  | .692  |
| Overall (50 items)  | 50        | .453   | .160  | .715  |

Table 4 reports the inter-scale (composite-score) correlation matrix. All five subscales correlated strongly and positively with one another ( $r = .68$ – $.82$ ), with Evaluating and Responding showing the strongest association ( $r = .820$ ) and Receiving/Attending and Evaluating the weakest ( $r = .677$ ) — still a large effect by conventional standards. These uniformly strong positive correlations are consistent with a single, strong underlying “general listening” dimension running through all five hypothesized stages, foreshadowing the confirmatory and structural results reported next.

Table 4. Inter-Scale (Composite Score) Correlation Matrix

| Subscale            | Receiving/Attending | Understanding | Remembering | Evaluating | Responding |
|---------------------|---------------------|---------------|-------------|------------|------------|
| Receiving/Attending | —                   | .810          | .728        | .677       | .730       |
| Understanding       | .810                | —             | .751        | .750       | .736       |
| Remembering         | .728                | .751          | —           | .742       | .747       |
| Evaluating          | .677                | .750          | .742        | —          | .820       |
| Responding          | .730                | .736          | .747        | .820       | —          |

#### 4.4 Confirmatory Factor Analysis

To formally test the measurement structure suggested by the EFA, four nested models were estimated by maximum likelihood on the screened sample ( $N = 291$ ) using semopy: (1) the a priori five-factor correlated model corresponding to the original subscale design; (2) an EFA-informed five-factor correlated model using the empirical item groupings in Table 2 (cross-loading items assigned to their highest-loading factor); (3) a second-order (hierarchical) model in which the five a priori first-order factors load onto a single higher-order General Listening Process factor; and (4) a full structural model extending Model 3 with a structural path from the General factor to a numeric approximation of academic performance (bracket midpoints: below 80 = 75, 80–85 = 82.5, 85–89 = 87, 90–100 = 95). Fit indices for all four models are reported in Table 5.

Table 5. Fit Indices for Confirmatory and Structural Models

| Model                                          | df   | $\chi^2$ | p     | CFI  | TLI  | RMSEA | AIC    |
|------------------------------------------------|------|----------|-------|------|------|-------|--------|
| Model 1: A priori 5-factor (correlated)        | 1165 | 2240.08  | <.001 | .883 | .877 | .056  | 204.60 |
| Model 2: EFA-informed 5-factor (correlated)    | 1165 | 2326.55  | <.001 | .873 | .867 | .059  | 204.01 |
| Model 3: Second-order (hierarchical) factor    | 1170 | 2305.35  | <.001 | .876 | .870 | .058  | 194.16 |
| Model 4: Full structural model (+ GWA outcome) | 1219 | 2371.69  | <.001 | .874 | .869 | .057  | 197.70 |

All four models produced RMSEA values in the “acceptable” range (.056–.059, below the .06–.08 conventional ceiling) but CFI and TLI values below the conventional .90–.95 thresholds (.873–.883), indicating adequate but not excellent absolute fit — an outcome broadly consistent with the merged and split factors already identified in the EFA. The a priori five-factor model (Model 1) showed marginally better fit by chi-square, CFI, and TLI than the EFA-informed model (Model 2), while the second-order model (Model 3) fit nearly identically to Model 1 despite being more parsimonious (5 additional degrees of freedom for essentially the same chi-square) and achieved the lowest AIC of the three measurement models. This pattern — a hierarchical model matching the fit of a freely correlated first-order model at lower complexity — is itself evidence for the strong general listening factor: standardized loadings of the five first-order factors onto the second-order General factor were uniformly high, ranging from .880 (Evaluating) to .912 (Understanding), indicating that all five hypothesized stages are strongly saturated by a common underlying listening competence. Standardized item loadings on their respective first-order factors in the a priori model ranged from .54 to .81 ( $M = .71$ ), all well above the conventional .40 minimum for retention (Watkins, 2018).

#### 4.5 Structural Equation Modeling: Listening Process and Academic Performance

Model 4 extended the second-order measurement model with a structural regression path from the General Listening Process factor to students’ academic performance (operationalized as the numeric midpoint of their reported general-average bracket). This path was positive and statistically significant (unstandardized  $B = 2.27$ , standardized  $\beta = .212$ ,  $z = 3.38$ ,  $p < .001$ ), indicating that a one-standard-deviation increase in the latent General listening factor was associated with roughly a 0.21-standard-deviation increase in academic performance, holding the measurement model constant. Overall model fit for the full structural model (Table 5, Model 4) was essentially unchanged from the pure measurement model (Model 3), confirming that the structural addition did not degrade fit. This structural result corroborates, within a full latent-variable framework, the nonparametric known-groups finding reported in Section 4.8 (Table 8) and provides the strongest single piece of evidence in this study that the LPS’s General factor captures a construct with genuine academic relevance rather than a measurement artifact.

#### 4.6 Internal Consistency Reliability

Cronbach’s alpha for the full 50-item scale was .976 ( $N = 314$ ), and alpha values for the five a priori subscales ranged from .900 to .937 (Table 6), all exceeding the conventional .80 threshold for research instruments (Taber, 2018). Corrected item-total correlations exceeded .30 for every item, and no subscale’s alpha would have improved materially through removal of any single item, indicating that all 50 items contribute meaningfully to their respective subscales despite the more nuanced structure revealed by the factor analysis.

Table 6. Internal Consistency Reliability of the LPS

| Subscale            | k (items) | Cronbach’s $\alpha$ | Interpretation |
|---------------------|-----------|---------------------|----------------|
| Receiving/Attending | 10        | 0.900               | Excellent      |
| Understanding       | 10        | 0.914               | Excellent      |
| Remembering         | 10        | 0.913               | Excellent      |
| Evaluating          | 10        | 0.937               | Excellent      |
| Responding          | 10        | 0.922               | Excellent      |
| Overall Scale       | 50        | 0.976               | Excellent      |

#### 4.7 Descriptive Level of the Listening Process

Using the conventional five-point verbal-interpretation range (1.00–1.79 Very Low, 1.80–2.59 Low, 2.60–3.39 Moderate, 3.40–4.19 High, 4.20–5.00 Very High), respondents rated themselves at a “High” level on all five listening dimensions and overall (Table 7), with means ranging narrowly from 3.47 (Remembering) to 3.67 (Responding). The Remembering dimension had both the lowest mean and among the more restricted subscale reliabilities, consistent with retention being commonly reported as the most effortful stage of listening in prior literature (Wolvin & Coakley, 1996).

Table 7. Descriptive Levels of the Listening Process

| Dimension                 | Mean | SD   | Verbal Interpretation |
|---------------------------|------|------|-----------------------|
| Receiving/Attending       | 3.66 | 0.67 | High                  |
| Understanding             | 3.55 | 0.64 | High                  |
| Remembering               | 3.47 | 0.63 | High                  |
| Evaluating                | 3.60 | 0.66 | High                  |
| Responding                | 3.67 | 0.67 | High                  |
| Overall Listening Process | 3.59 | 0.58 | High                  |

#### 4.8 Differences Across Profile Variables

Because Shapiro–Wilk testing indicated the overall composite score departed significantly from normality ( $W = .977$ ,  $p < .001$ ), nonparametric tests were used for group comparisons (Table 8). Female respondents reported significantly higher scores than male respondents on Understanding (Mdn difference,  $p = .049$ ), Remembering ( $p = .045$ ), Responding ( $p = .017$ ), and the Overall composite ( $p = .029$ ), though not on Receiving/Attending or Evaluating. Neither age bracket nor first-language background produced statistically significant differences on any dimension. Academic performance bracket, by contrast, was significantly associated with every listening dimension (all  $p \leq .007$ ), with mean Overall LPS scores rising monotonically from 3.28 among students in the below-80 bracket to 3.51 (80–85), 3.51 (85–89), and 3.79 (90–100).

Table 8. Group Comparisons Across Profile Variables

| Dimension           | Sex (U)  | Sex (p) | Age (H)    | Age (p) | Language | GWA      | GWA (p) | p-value |
|---------------------|----------|---------|------------|---------|----------|----------|---------|---------|
| Receiving/Attending | 10339.50 | .116    | 13.19 (ns) | .528    | not sig. | not sig. | 12.27   | .007*   |
| Understanding       | 10031.50 | .049*   | 15.20 (ns) | .604    | not sig. | not sig. | 14.60   | .002*   |
| Remembering         | 10009.00 | .045*   | 16.63 (ns) | .408    | not sig. | not sig. | 21.00   | <.001*  |
| Evaluating          | 10693.50 | .261    | 14.19 (ns) | .908    | not sig. | not sig. | 14.39   | .002*   |
| Responding          | 9711.00  | .017*   | 13.16 (ns) | .923    | not sig. | not sig. | 12.07   | .007*   |
| Overall             | 9859.00  | .029*   | 13.74 (ns) | .691    | not sig. | not sig. | 17.09   | .001*   |

This monotonic association between self-reported listening scores and academic performance provides meaningful criterion-related validity evidence: a scale capturing a construct genuinely relevant to classroom learning would be expected to covary with an independent academic outcome, and it does so consistently across every hypothesized dimension of the instrument.

## 5. Discussion

This study set out to validate a stage-based measure of the listening process for HEI students in Region II, Philippines, and to describe the level at which students engage in each stage. Four findings merit discussion.

First, the LPS demonstrated excellent internal consistency at both the subscale and full-scale level, and excellent sampling adequacy for factor analysis — indicating that the item pool is coherent and well-suited to measuring self-reported listening behavior in this population. The very high overall alpha (.976), however, should be interpreted cautiously rather than celebrated uncritically: alpha values in this range can reflect item redundancy as much as true reliability (Taber, 2018), and the strong general factor identified in the parallel analysis is

consistent with this interpretation. Future revisions of the LPS might consider trimming redundant items within each subscale without materially reducing reliability, thereby producing a shorter, more efficient instrument.

Second, the empirical factor structure diverged from the a priori five-stage design in an interpretable way. The receiving/attending and understanding stages merged into a single Focused Reception factor, while evaluating and remembering emerged as clean, distinct factors, and responding split into two related facets. This pattern echoes a recurring finding in the broader listening-measurement literature, where self-report listening instruments tend to produce a dominant general factor alongside a smaller number of more specific facets rather than cleanly separable stage-by-stage factors (Bodie, Worthington, Imhof, & Cooper, 2008; Bodie, 2013). One plausible explanation is that receiving and understanding occur nearly simultaneously during real-time listening, making them difficult for respondents to introspectively disentangle when rating their own behavior, whereas remembering (a later, more effortful and consciously monitored process) and evaluating (a more deliberate, judgment-based process) are more psychologically distinguishable and therefore easier to rate as separate constructs. This suggests that instrument developers targeting the early stages of the listening process may need either more behaviorally granular items or a different measurement approach (e.g., performance-based rather than self-report tasks) to achieve empirical separation.

Third, the confirmatory and structural analyses converge with, and extend, the exploratory findings. None of the four CFA/SEM models reached conventional “good fit” thresholds on CFI or TLI, even though RMSEA fell in the acceptable range for all of them — a common pattern for long, highly intercorrelated Likert scales, and one that should temper any claim that the LPS’s factor structure is fully confirmed. More informative than any single model’s fit, however, is the comparison across models: a second-order model in which all five stages loaded onto one General Listening Process factor matched the fit of the more complex, freely correlated first-order model at a lower AIC, and every first-order factor loaded on that general factor at .88 or above. Read together with the uniformly high inter-scale correlations (Table 4) and the elevated inter-item correlations within every subscale (Table 3), this is convergent, multi-method evidence — from EFA, CFA, and inter-item correlation analysis alike — that respondents’ self-reported listening behavior is best described as one dominant listening-competence dimension expressed through five correlated facets, rather than five fully independent abilities. This has a direct practical implication: institutions adopting the LPS should treat the Overall/General score as the primary, best-supported unit of measurement, using the five subscale scores as diagnostic supplements rather than as independently validated stand-alone measures until further refinement reduces item redundancy and cross-loading.

Fourth, the known-groups and structural analyses offer encouraging, if partial, validity evidence. The monotonic relationship between academic performance bracket and every LPS dimension — corroborated by the significant structural path from the latent General factor to academic performance in the full SEM ( $\beta = .212, p < .001$ ) — is the strongest evidence in this study that the scale captures something practically consequential rather than an artifact of item wording. The sex difference — female respondents scoring higher on three of five dimensions and overall — is consistent with prior communication research reporting modestly higher self-reported listening or interpersonal-communication competence among female respondents (Bodie, 2011), though self-report differences of this kind may also reflect response-style differences between groups rather than differences in underlying listening skill, a distinction this design cannot resolve. The absence of differences by age bracket or first-language background suggests that, within this sample, the listening process as measured by the LPS operates comparably across these strata.

Practically, these findings suggest that HEI faculty and student-support units in Region II can use the LPS as a diagnostic screening tool, particularly for identifying students who may benefit from targeted support in retention (Remembering) or critical evaluation, the two stages that emerged as the cleanest, most distinguishable empirical dimensions and therefore the most defensible for individual-level score interpretation. Interpreting Receiving/Attending and Understanding as a single “Focused Reception” composite, rather than as two separate scores, is recommended until confirmatory work establishes otherwise.

## 6. Limitations

Several limitations qualify these findings. First, the sample was recruited through convenience, online sampling rather than probability sampling, which limits generalizability to the broader population of Region II HEI students and may have contributed to the data-quality concerns (23 straight-lining respondents) documented in Section 3.4. Second, all data are self-reported; the LPS measures students’ perceptions of their own listening behavior, which may diverge from observed or performance-based listening competence. Third, although this study went beyond exploratory factor analysis to test the measurement structure with confirmatory factor analysis and a structural equation model, all four CFA/SEM models were estimated on the same sample used to inform the EFA-based model (Model 2) and, for Model 1, the same sample used to design the a priori subscales; independent-sample cross-validation, ideally with a second Region II or national cohort, is needed before

treating the confirmatory fit indices as fully out-of-sample evidence. Relatedly, the ratio of sample size ( $N = 291$ ) to freely estimated parameters in the 50-item CFA/SEM models (roughly 2.5:1) falls well short of common recommendations of 10:1 or higher, which can make chi-square and standard errors less trustworthy even though CFI, TLI, and RMSEA are comparatively robust to this concern; the fit indices reported here should accordingly be treated as provisional pending replication on a substantially larger sample. Fourth, test-retest reliability and convergent validity against an established listening instrument (e.g., the Watson–Barker Listening Test) were not assessed in this study and represent a priority for follow-up validation work. Fifth, the numeric academic-performance variable used in the structural model was approximated from categorical general-average brackets using bracket midpoints, a coarse proxy for actual grade point average that likely attenuates the true association between listening and academic performance. Finally, the cross-sectional, correlational design of the known-groups and structural analyses cannot establish that stronger listening skills cause better academic performance; the association is consistent with, but does not confirm, a causal relationship in either direction.

## 7. Conclusion and Recommendations

The Listening Process Scale demonstrated excellent internal consistency, strong sampling adequacy, and a coherent, hierarchically-organized factor structure — substantiated across exploratory factor analysis, inter-item/inter-scale correlation analysis, confirmatory factor analysis, and structural equation modeling — when validated on a sample of 314 (291 after data-quality screening) HEI students in Region II, Philippines. Students reported engaging in all five stages of the listening process at a consistently “High” level, and a latent General Listening Process factor showed a significant structural path to academic performance, supporting the instrument’s practical relevance beyond simple mean-level description. The instrument’s Evaluating and Remembering subscales performed as the most empirically distinct within the EFA and are recommended as diagnostic scores for identifying stage-specific weaknesses; the Receiving/Attending and Understanding items are recommended for interpretation as a single composite pending further refinement; and the Responding items may benefit from a formal two-facet revision distinguishing embodied engagement cues from verbal follow-up behavior. Because none of the CFA/SEM models reached conventional good-fit thresholds on CFI/TLI, and because the strong second-order general factor indicates considerable shared variance across all five stages, the single Overall/General score is currently the most defensible unit for high-stakes or comparative use, with the five subscale scores serving a formative, diagnostic role.

Future research should pursue confirmatory factor analysis and structural equation modeling on an independent Region II or nationally representative sample — ideally with a larger  $N$  to raise the parameter-to-sample ratio beyond that available here — establish convergent validity against an existing listening measure, assess test-retest reliability, and explore measurement invariance across sex and institution type. Item reduction targeting the six flagged cross-loading items (Q2, Q12, Q13, Q18, Q29, Q30) is a natural next step toward a shorter instrument with cleaner simple structure. Researchers and institutions adopting the LPS are encouraged to retain the data-quality screening procedures documented here, given the meaningful proportion of straight-lining responses observed in this online-administered sample.

## References

- Agucito, R. G., Macatabon, R. A., & Imbang-Agni, E. J. D. (2025). Factors affecting the listening skills of college students. *E-DAWA*, 5(2), 1–15. University of Southern Mindanao.
- Bodie, G. D. (2011). The Active-Empathic Listening Scale (AELS): Conceptualization and evidence of validity within the interpersonal domain. *Communication Quarterly*, 59(3), 277–295. <https://doi.org/10.1080/01463373.2011.583495>
- Bodie, G. D. (2013). Issues in the measurement of listening. *Communication Research Reports*, 30(1), 76–84. <https://doi.org/10.1080/08824096.2012.733981>
- Bodie, G. D., & Worthington, D. L. (2010). Revisiting the Listening Styles Profile (LSP-16): A confirmatory factor analytic approach to scale validation and reliability estimation. *International Journal of Listening*, 24(2), 69–88. <https://doi.org/10.1080/10904011003744516>
- Bodie, G. D., Worthington, D. L., & Fitch-Hauser, M. (2011). A comparison of four measurement models for the Watson–Barker Listening Test (WBLT)-Form C. *Communication Research Reports*, 28(1), 32–42. <https://doi.org/10.1080/08824096.2011.540547>
- Bodie, G. D., Worthington, D. L., & Gearhart, C. C. (2013). The Listening Styles Profile-Revised (LSP-R): A scale revision and evidence for validity. *Communication Quarterly*, 61(1), 75–93. <https://doi.org/10.1080/01463373.2012.720343>
- Bodie, G. D., Worthington, D. L., Imhof, M., & Cooper, L. O. (2008). What would a unified field of listening look like? A proposal linking past perspectives and future endeavors. *International Journal of Listening*,

- 22(2), 103–122. <https://doi.org/10.1080/10904010802174867>
- Brownell, J. (1985). A model for listening instruction: Management applications. *Bulletin of the Association for Business Communication*, 48(3), 39–44.
- Brownell, J. (1996). *Listening: Attitudes, principles, and skills*. Allyn & Bacon.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309–319. <https://doi.org/10.1037/1040-3590.7.3.309>
- Gilakjani, A. P., & Ahmadi, M. R. (2011). A study of factors affecting EFL learners' English listening comprehension and the strategies for improvement. *Journal of Language Teaching and Research*, 2(5), 977–988.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Igolkina, A. A., & Meshcheryakov, G. (2020). semopy: A Python package for structural equation modeling. *Structural Equation Modeling*, 27(6), 952–963. <https://doi.org/10.1080/10705511.2019.1704289>
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455. <https://doi.org/10.1037/a0028085>
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, 44(3), 219–246. <https://doi.org/10.1177/0095798418771807>
- Wolvin, A. D., & Coakley, C. G. (1996). *Listening* (5th ed.). Brown & Benchmark.

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

# Discipline or Pressure? Examining the Effectiveness of the KDM Military Barracks in Developing Students' Learning Motivation

Desti Agustin<sup>1</sup>

<sup>1</sup> Faculty of Psychology, Universitas Airlangga, Surabaya, Indonesia

Correspondence: Desti Agustin, Faculty of Psychology, Universitas Airlangga, Surabaya, Indonesia.

doi:10.63593/RAE.2788-7057.2026.06.005

## Abstract

The phenomenon of juvenile delinquency in West Java, Indonesia, has triggered a unique policy response from Governor Dedi Mulyadi, namely a rehabilitation program through military barracks for students involved in deviant behavior. This article aims to analyze the effectiveness of this approach in developing students' learning motivation, utilizing a theoretical review based on Self-Determination Theory, Attribution Theory, and Self-Efficacy Theory. The study was conducted employing a qualitative-descriptive method with narrative and reflective analysis of secondary data, encompassing media reports, public policies, and educational psychology literature. The findings indicate that although the military barracks program may yield short-term behavioral changes, its long-term success heavily relies on the integration of reflective strategies, psychological support, and social reintegration. If managed sensitively and sustainably, military barracks have the potential to serve as an initial step in building students' intrinsic motivation and positive character. However, without fulfilling basic psychological needs such as autonomy and relatedness, this approach risks creating counterproductive pressure on students' learning and emotional well-being.

**Keywords:** learning motivation, military barracks, juvenile delinquency, educational psychology, reflective approach, West Java, Indonesia

## 1. Introduction

The phenomenon of juvenile delinquency in West Java has become a serious concern in recent years. Although data from the Central Bureau of Statistics (BPS) indicates a decline in the number of cases from 12,345 in 2020 to 10,890 in 2022, this figure still reflects a significant challenge within the educational and social spheres (Kemen PPPA, 2025). Common forms of delinquency include bullying, drug abuse, brawls, motorcycle gangs, and truancy (Antaranews, 2025). In response to this situation, the Governor of West Java, Dedi Mulyadi, initiated a rehabilitation program for troubled students utilizing a military barracks approach (Kompas, 2025b).

This program aims to instill the values of discipline, responsibility, and integrity in students considered "deviant" or who have lost direction in their education (Kompas, 2025b). While this approach has garnered public support a survey by Indikator Politik Indonesia noted that 92% of respondents support the policy (CNN Indonesia, 2025) it has also drawn criticism from various parties. The Indonesian Child Protection Commission (KPAI), for instance, has highlighted potential violations of children's rights during its implementation (Kompas, 2025a).

In the realm of education, learning motivation is a crucial factor that heavily influences students' academic success. Various efforts have been made to enhance this motivation, including innovative approaches such as the one implemented by Kang Dedi Mulyadi (KDM) through the military barracks program. This program seeks to integrate the strict discipline typically enforced in the military with formal education, with the hope of boosting students' learning motivation. However, the ensuing question is whether this approach functions more as constructive discipline or rather as detrimental pressure?

According to research by Bai et al. (2022), there is a significant relationship between physical exercise and students' academic adaptation, where regular physical activity can foster a stronger enthusiasm for learning. In this context, the KDM military barracks program—which integrates physical and mental components—often succeeds in building an environment that supports students' emotional and cognitive development. The physical facilities and disciplined routines within the barracks are expected to assist students in developing self-management skills, which concurrently enhance their learning motivation.

Nevertheless, it is undeniable that there are inherent risks associated with this approach, particularly regarding the level of pressure perceived by the students. Although the study by Ahern et al. (2015) focuses on the challenges of veterans returning from military duty rather than students in military barracks, the insights regarding the difficulties in finding structure and purpose in life can reflect the challenges students might face if the pressure is not managed properly. If left unmanaged, the applied discipline can act as pressure, leading to high stress levels that could negatively impact their mental health and learning motivation.

Striking a balance between discipline and emotional support is vital in this process. Further research is necessary to explore how the environmental conditions in the barracks can affect both the physical and mental well-being of the students. In this analysis, the key to improving learning motivation within the KDM military barracks lies in the ability to manage strict discipline without compromising the students' emotional well-being. This program must be designed not only to encourage discipline but also to provide space for students' exploration, creativity, and self-development. This is essential to ensure that educational goals are not solely focused on compliance, but also on character development and critical thinking skills, enabling students to achieve better academic performance.

This phenomenon has inevitably captured public attention. Many have praised KDM's actions as a much-needed "wake-up call" for today's students. However, many others question: does a militaristic approach genuinely succeed in developing students' learning motivation, or does it merely create momentary compliance rooted in fear?

Answering this question through the lens of educational psychology, rather than mere moral opinion, is crucial. Learning motivation is not simply the product of commands or punishments; rather, it stems from feeling capable, accepted, and having control over the learning process. Therefore, this article endeavors to examine the KDM military barracks approach through learning motivation theories such as Self-Determination Theory, Attribution Theory, and Self-Efficacy Theory. Additionally, the roles of the learning climate and the experience of engagement (flow) in sustaining post-rehabilitation motivation will be discussed.

Can harsh discipline give rise to true motivation? Or are we repeating an educational pattern that punishes before it understands? This article invites readers to look deeper, beyond the virality of the action, toward a more reflective and theory-based understanding. Ultimately, this study aims to analyze the military barracks approach initiated by Dedi Mulyadi from the perspective of learning motivation theories. By examining empirical data and educational psychology theories, this research hopes to provide a more profound understanding of whether this approach is truly effective in building students' intrinsic motivation or if it only yields temporary behavioral changes driven by external pressure.

## **2. Methods**

This article was structured employing a qualitative-descriptive approach alongside narrative and theoretical analysis methods (Creswell, 2015). It is based on a case study of the policy concerning the rehabilitation of troubled youth through military barracks, initiated by Kang Dedi Mulyadi (KDM) in West Java. The nature of this paper is a scientifically-backed academic opinion piece (Riessman, 2008), wherein data were collected and analyzed from the following three primary sources:

### **1) Secondary Data from Online Media**

This includes news articles, interviews, and social media documentation published between 2023 and 2025 that cover the student rehabilitation activities in the military barracks by KDM. The media sources referenced include prominent national outlets such as Antaranews, Kompas, and CNN Indonesia.

### **2) Educational Psychology Theoretical Review**

This article grounds its analysis in highly relevant learning motivation theories, specifically: Self-Determination Theory (Deci & Ryan, 2000), Attribution Theory (Weiner, 1986), and Self-Efficacy Theory (Bandura, 1997).

### **3) Conceptual and Reflective Analysis**

The authors synthesized practical phenomena in the field with theoretical frameworks through a reflective analysis, without involving direct research subjects. This approach facilitates a critical examination of educational policies by thoroughly considering the psychological and social factors affecting students.

This article does not utilize primary data, such as field interviews or direct observations. Instead, it aims to present a robust academic perspective designed to serve as a foundational basis for policy reflection and broader educational discourse.

### 3. Results and Discussion

#### 3.1 Background: Juvenile Delinquency in West Java

Juvenile delinquency is no longer merely a marginal issue; it has evolved into an alarming social phenomenon, particularly in heavily populated provinces like West Java. With a large student population and rapid urbanization dynamics, this region has become an epicenter for deviant behavior among adolescents. This phenomenon does not occur in isolation but is the result of overlapping factors such as weak social control, family fragmentation, peer pressure, and unlimited access to digital information without adequate literacy.

Factual data reinforces this urgency. The Sukabumi Police Department recorded an increase in student brawls and the use of sharp weapons by adolescents in early 2025, indicating that physical violence is no longer a sporadic issue but a deeply rooted pattern. In the Tasikmalaya area, SMPN 3 Ciawi deemed it necessary to collaborate with the police to provide direct education to students as a preventive measure against the rising trend of youth violence (Antaranews, 2025). Furthermore, according to a report by the Ministry of Women's Empowerment and Child Protection, as of January 1, 2024, there were 24,197 recorded cases of violence against children, with females constituting the majority of victims (Kemen PPPA, 2025). These figures indicate that the emotional and social conditions of children and adolescents in Indonesia are under serious structural pressure.

Juvenile delinquency encompasses not only legal violations or physical violence but also patterns of disengagement from the learning process, low motivation, truancy, and disrupted social relationships, which ultimately lead to an inability to form positive identities as productive individuals. This aligns with Sulaiman's (2022) study in *Remaja: Antara Kemandirian dan Kenakalan*, which states that delinquency does not merely emerge from within the individual but reflects the dysfunction of the social system in which the adolescent develops. Sulaiman emphasizes that the neglect of adolescents' affective needs in families and schools is the root cause that exacerbates deviant behavior.

Approaches that rely solely on punishment or forced discipline have proven insufficient. Various studies indicate that interventions ignoring the psychosocial needs of adolescents risk causing long-term impacts, including anxiety, self-doubt, and an inability to form intrinsic motivation. Therefore, it is crucial to formulate a response that is strategic and integrated, rather than merely reactive to the symptoms of delinquency.

Against this backdrop, the military barracks policy for troubled students initiated by Kang Dedi Mulyadi becomes highly relevant for examination. Does this policy answer the adolescents' need for firm yet empathetic guidance? Or is it merely a new form of social control that neglects the importance of self-reflection and psychological recovery? Answering this requires an analysis that is not only descriptive but also theoretical, grounded in a deep understanding of learning motivation dynamics and adolescent psychosocial development.

#### 3.2 Government Response: The KDM Military Barracks Program

In response to the rising rates of juvenile delinquency in West Java, Governor Dedi Mulyadi introduced a controversial yet innovative policy: a military barracks program for troubled students. This program aims to restore the social function and moral responsibility of adolescents through a disciplinary approach based on military training. Students involved in deviant behaviors—such as brawls, illegal racing, motorcycle gangs, and insubordination to school authority—are gathered and sent to a barracks facility to undergo a strict routine.

Activities in the barracks are not limited to physical training and marching but also include motivational lectures, mental conditioning, and daily disciplines such as waking up early, communal work, and a temporary disconnection from the digital world. The primary goal is to create a “shock therapy” effect that encourages adolescents to reflect on their actions, realize the social consequences of their deviant behavior, and reshape their mindsets and personal characters.

From a policy perspective, this step represents a formal social control intervention, where the state assumes a direct role in handling adolescent issues previously considered the domain of the family or school. Dedi Mulyadi asserts that character education cannot be left entirely to the formal curriculum; it must be safeguarded through concrete actions that touch upon behavioral awareness at the community level.

However, the effectiveness of this approach cannot be assessed solely by the emotional responses of students who cry upon returning home or promise not to repeat their mistakes. A deeper understanding is required regarding the internalization of values, the sustainability of learning motivation, and character resilience after students return to their original environments.

A study by Moeller et al. (2020) indicates that physical discipline-based rehabilitation programs can reduce deviant behavior among adolescents, provided they are not solely punishment-oriented but incorporate

supportive approaches, psychological recovery, and relational development. If such programs fail to provide space for personal reflection, safe emotional engagement, and long-term social support, the resulting behavioral changes will be superficial and short-lived.

Furthermore, from the perspective of social learning theory (Bandura), military barracks approaches have two sides. On one hand, models of disciplined behavior and consistency can serve as powerful observational learning, especially if authoritative figures in the barracks demonstrate a combination of firmness and empathy. On the other hand, if the students' emotional management is neglected, there is a risk of passive resistance or trauma toward authority, which will further alienate students from meaningful learning processes.

To be systemically successful, this military barracks program must be situated within a sustainable character education framework, emphasizing restorative practices rather than mere behavioral correction. This means that barracks rehabilitation must be followed by Reintegration support at school, Post-rehabilitation counseling and monitoring, Collaboration with parents and school counselors, and Regular evaluations of students' motivation and self-regulation.

With this framework, the military barracks program can serve as an initial step, rather than the final step, in transforming students' learning motivation. Successful interventions are those that make students want to learn out of awareness, not merely comply out of fear.

### *3.3 Theoretical Analysis: The Program's Impact on Learning Motivation*

The military barracks approach applied by Dedi Mulyadi to troubled students is compelling for an in-depth analysis from an educational psychology perspective, particularly regarding the development of learning motivation. Motivation, as a psychological process that drives engagement, persistence, and goal orientation in learning, is not solely influenced by the environment, but also by students' perceptions of control, meaning, and achievement. Therefore, the effectiveness of this program cannot be adequately assessed merely through short-term behavioral changes, but rather through the extent to which it is capable of influencing learning motivation internally and sustainably.

#### *3.3.1 Self-Determination Theory*

According to Deci & Ryan's (2000) Self-Determination Theory (SDT), healthy and sustainable learning motivation develops when three basic psychological needs are fulfilled: autonomy, competence, and relatedness.

**Autonomy:** In a military barracks, this need is most vulnerable to being thwarted. Students do not have the choice to participate in the program, and all activities are based on strict instructions. When individuals feel they lack control over their lives, the resulting motivation tends to be extrinsically controlled—that is, learning or changing due to external pressure rather than internal drive or personal understanding.

**Competence:** Although the barracks environment is structured and high-pressure, it also creates concrete challenges that, if successfully navigated, can trigger a sense of self-confidence and success (mastery experience). Students who consistently manage to wake up early, participate in physical training, and complete barracks assignments may begin to perceive themselves as “capable” individuals, which has the potential to reinforce positive motivation—provided, of course, that this experience is internalized as the result of personal effort rather than coercion.

**Relatedness:** Amidst physical and emotional pressure, solidarity among students and relationships with instructors have the potential to create a sense of relatedness. If the relationships formed are supportive and empathetic, students can feel emotional support that strengthens their desire to change. However, if those relationships are based on fear or domination, the effect is precisely counterproductive to motivation.

Thus, while the potential for the barracks program to build new motivation does exist, it will only be optimal if there are conscious efforts to restore students' autonomy through reflection, providing a voice, and acknowledging changes that emerge from within the students themselves.

#### *3.3.2 Attribution Theory*

Attribution, according to Weiner (1986), is the way individuals explain the causes of their successes or failures. Weiner elucidates that healthy attributions are internal, stable, and controllable (for example, “I succeeded because I worked hard”). Unfortunately, in the context of military barracks, students who are publicly humiliated or forced to undergo harsh discipline may develop flawed attribution patterns:

*“I changed because I was afraid of being punished.”*

*“I only comply because I am being watched.”*

Such attribution patterns are dangerous because they:

*Do not encourage personal responsibility for the change.*

*Instill the idea that self-control depends on an authoritative figure.*

*Cause students to revert to old behaviors once the external pressure is removed.*

To prevent this, the rehabilitation approach must be accompanied by reflective exercises, group discussions, and recognition of the students' personal efforts—so that the attribution of success shifts toward being internal and stable.

### 3.3.3 Self-Efficacy Theory

Self-efficacy, according to Bandura (1997), is an individual's belief that they are capable of performing an action to achieve a specific goal. In the barracks program, strenuous physical activities, strict discipline, and daily achievements have the potential to create a mastery experience, which is the direct experience of knowing, "I can do this."

However, this positive effect only emerges if (1) The challenges provided fall within the students' proximal zone of development (neither too easy nor too difficult). (2) There is verbal support from a trusted figure (e.g., an instructor or mentor). (3) Students are given the opportunity to reflect on their achievements as the result of personal effort, not merely the outcome of pressure.

Without these conditions, the barracks program could actually undermine students' self-efficacy. They might feel that change is only possible under coercion, rather than through their own inner strength. This is highly dangerous, as low self-efficacy strongly correlates with a decline in learning motivation, increased anxiety, and withdrawal from academic challenges.

### 3.3.4 Theoretical Integration and Potential Risks

These three theories synergistically underscore that healthy learning motivation cannot be forged through discipline or pressure alone. Motivation that endures and has an impact requires:

- (1) Meaningful experiences that are internalized.*
- (2) A sense of having control over the change.*
- (3) Empathetic, rather than oppressive, social support.*

The military barracks program has the potential to serve as an "initial bridge" to interrupt students' destructive patterns. However, without a reflective and continuous psychological approach, the motivation formed will remain superficial, fragile, and quick to evaporate once the oppressive environment disappears.

### 3.4 Reflection and Policy Implications

The military barracks program initiated by Kang Dedi Mulyadi indeed demonstrates potential as an early intervention for students trapped in juvenile delinquency who have lost their learning motivation. The shock effect of strict discipline, physical routines, and restricted access to the outside world provides space for students to experience a "break" from their destructive lifestyles. In the short term, this approach can reinstall the value of responsibility, halt deviant behavior, and create a momentum for behavioral change. As an intervention strategy, it is crucial to adopt a more reflective approach that encourages self-awareness and character development. Research supporting this indicates that reflection in the learning process not only assists in developing academic skills but also in fostering students' social and emotional awareness (Ryan & Ryan, 2013). In higher education, the theme of reflection has been integrated into the curriculum, implying the need for a more comprehensive handling of troubled students.

However, the immediate changes apparent post-barracks do not necessarily reflect a profound motivational transformation. Based on Self-Determination Theory, healthy motivation does not grow from fear, shame, or social pressure, but rather from personal awareness, self-confidence, and success experiences that are internally constructed. Therefore, if students are merely controlled through discipline and are not equipped with the tools to understand and manage themselves, the motivation will be superficial and unsustainable.

The policy implication is that the barracks must not stand alone as a singular solution. It must be situated within a multi-layered and sustainable intervention framework. First, a structured reintegration into the school environment is necessary, where teachers, homeroom advisors, and school counselors are actively involved to assist the students' readaptation process. Second, psychological support is mandatory, both in the form of individual and group counseling, to help students process the barracks experience reflectively rather than traumatically.

To achieve optimal results, the involvement of various parties, including schools, guidance counselors, and psychological support, is paramount. This aligns with findings demonstrating that family involvement in child development, particularly in complex interventions, can enhance the effectiveness of educational programs (Sousa et al., 2023). Furthermore, continuous psychological support is also essential for building self-efficacy,

namely the students' belief in their ability to succeed in specific situations and their intrinsic motivation (Boe et al., 2018). This process helps students feel more engaged and responsible for their learning and personal development.

In the development of character and identity, the military barracks program must be complemented by an approach oriented toward emotional well-being. This is because, although physical training and discipline are vital components of military programs, the psychological impact of such training is often overlooked (Crowley et al., 2015). Research shows that the formation of a strong self-identity among military students can improve their competitiveness and commitment to the organization (Johansen et al., 2014). By integrating training that supports reflection and character development, the program can deliver a deeper and more long-term impact on the students.

Moreover, students must be positioned not as "objects of a disciplinary project," but as individuals with complex backgrounds, emotional needs, and the potential to grow. A humanistic approach based on relationships and understanding, rather than punishment, is more effective in shaping long-term learning motivation. Additionally, family involvement is crucial so that the values built in the barracks are not disconnected when students return to their original environments.

Furthermore, establishing clear and measurable success targets according to inclusive educational policies is crucial. By paying attention to social and emotional factors in training, this program can not only mitigate juvenile delinquency behaviors but also increase the likelihood of students reintegrating into society after completing the program. This is further supported by studies showing that a supportive environment and a deeper understanding of social dynamics can enrich students' learning experiences (Burks, 2011). Thus, it is highly recommended to create a more holistic intervention model that involves various stakeholders in the change process.

Finally, policies like military barracks require integration with an educational system that favors recovery, rather than mere correction. Education is fundamentally not just about fixing behavior, but also fostering self-belief, a sense of meaning, and a healthy relationship with the learning process. If these aspects can be synergized, the barracks approach could become the genesis of transformation, rather than just another form of pressure polished with an educational facade. The reflection and policy implications suggest that although the military barracks program may provide a good start for lost students, its long-term success will only be achieved through a process that focuses on emotional and social development, as well as the creation of an atmosphere that supports their reintegration into society.

#### 4. Conclusion

The military barracks program initiated by Kang Dedi Mulyadi in West Java represents a concrete response to the escalation of juvenile delinquency and the low learning motivation among students. Through a strict disciplinary approach, this program succeeds in delivering a shock effect, even fostering an initial awareness among some participants regarding the importance of returning to the educational track. Nevertheless, from the perspective of educational psychology, the long-term effectiveness of this approach warrants profound critical examination. Drawing upon Self-Determination Theory, Self-Efficacy Theory, and Attribution Theory, it becomes evident that the military barracks merely address a fraction of the aspects that constitute learning motivation, particularly within the dimensions of external control and competence. Without continuous mentoring and a supportive learning environment post-program, the ensuing changes risk being temporary and uninternalized. Motivation forged through pressure tends to be fragile and unsustainable unless it is transformed into intrinsic motivation through reflective processes, social support, and integration into the formal educational system. Thus, it is imperative to view approaches like the military barracks not as a solitary solution, but rather as an initial intervention that must be followed by profound pedagogical and psychological strategies.

#### 5. Recommendations

A follow-up approach more heavily oriented toward fulfilling the psychological needs of the students after their participation in the military barracks program is necessary. Local governments, schools, and families must synergize to create a learning environment that supports the growth of intrinsic motivation, rather than mere short-term compliance. Disciplinary interventions such as the barracks should serve as a gateway for a continuous rehabilitation process focused on self-reflection, the restoration of social relationships, and the inculcation of meaningful learning values.

#### References

- Ahern, J., Worthen, M., Masters, J., Lippman, S. A., Ozer, E. J., & Moos, R. (2015). The Challenges of Afghanistan and Iraq Veterans' Transition from Military to Civilian Life and Approaches to Reconnection. *PLOS ONE*, 10(7), e0128599. <https://doi.org/10.1371/journal.pone.0128599>

- Antaranews. (2025, February 8). Kenakalan remaja makin marak, SMPN 3 Ciawi menggandeng Polsek untuk edukasi. *Antaranews*.  
<https://radartasik.id/2025/02/08/kenakalan-remaja-makin-marak-smpn-3-ciawi-menggandeng-polsek-untuk-mencegah-dan-mengedukasi/>
- Bai, M.-Z., Yao, S.-J., Ma, Q.-S., Wang, X.-L., Liu, C., & Guo, K.-L. (2022). The relationship between physical exercise and school adaptation of junior students: A chain mediating model. *Frontiers in Psychology*, 13, 977663. <https://doi.org/10.3389/fpsyg.2022.977663>
- Bandura, A. (1997). *Self-efficacy: The exercise of control* (pp. ix, 604). W H Freeman/Times Books/ Henry Holt & Co.
- Boe, O., Säfvenbom, R., Johansen, R. B., & Buch, R. (2018). The Relationship between Self-Concept, Self-Efficacy, and Military Skills and Abilities. *International Journal of Learning, Teaching and Educational Research*, 17(10), 18–42. <https://doi.org/10.26803/ijlter.17.10.2>
- Burks, D. J. (2011). Lesbian, gay, and bisexual victimization in the military: An unintended consequence of “Don’t Ask, Don’t Tell”? *American Psychologist*, 66(7), 604–613. <https://doi.org/10.1037/a0024609>
- CNN Indonesia. (2025). *Survei: 92% Setuju Program Dedi Mulyadi Kirim Siswa ke Barak Militer*. Nasional. <https://www.cnnindonesia.com/nasional/20250528172047-20-1234266/survei-92-setuju-program-dedi-mulyadi-kirim-siswa-ke-barak-militer>
- Creswell, J. W. (2015). *Penelitian kualitatif dan desain riset: Memilih di antara lima pendekatan* (3rd ed.). Pustaka Pelajar.
- Crowley, S. K., Wilkinson, L. L., Wigfall, L. T., Reynolds, A. M., Muraca, S. T., Glover, S. H., Wooten, N. R., Sui, X., Beets, M. W., Durstine, J. L., Newman-Norlund, R. D., & Youngstedt, S. D. (2015). Physical Fitness and Depressive Symptoms during Army Basic Combat Training. *Medicine & Science in Sports & Exercise*, 47(1), 151–158. <https://doi.org/10.1249/MSS.0000000000000396>
- Deci, E. L., & Ryan, R. M. (2000). The “What” and “Why” of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, 11(4), 227–268. [https://doi.org/10.1207/S15327965PLI1104\\_01](https://doi.org/10.1207/S15327965PLI1104_01)
- Johansen, R. B., Laberg, J. C., & Martinussen, M. (2014). Military Identity as Predictor of Perceived Military Competence and Skills. *Armed Forces & Society*, 40(3), 521–543. <https://doi.org/10.1177/0095327X13478405>
- Kemen PPPA. (2025). *Kemen PPPA: Bersama Kawal Kebijakan Perlindungan dan Pemenuhan Hak Anak*. <https://www.kemenpppa.go.id/page/view/NTgzNw==>
- Kompas. (2025a). *KPAI Desak Dedi Mulyadi Setop Kirim Siswa ke Barak Militer Sampai Ada Evaluasi*. [https://nasional.kompas.com/read/2025/05/26/15083061/kpai-desak-dedi-mulyadi-setop-kirim-siswa-ke-barak-militer-sampai-ada?utm\\_source=chatgpt.com](https://nasional.kompas.com/read/2025/05/26/15083061/kpai-desak-dedi-mulyadi-setop-kirim-siswa-ke-barak-militer-sampai-ada?utm_source=chatgpt.com)
- Kompas. (2025b, March 5). *Dedi Mulyadi terapkan kurikulum wajib militer di SMA/SMK Jabar*. <https://bandung.kompas.com/read/2025/03/05/063144678/dedi-mulyadi-terapkan-kurikulum-wajib-militer-di-sma-smk-jabar-pada-tahun>
- Levy, J. W., Bhooniboonchoo, P., Simasathien, S., Salje, H., Huang, A., Rangsin, R., Jarman, R. G., Fernandez, S., Klungthong, C., Hussem, K., Gibbons, R. V., & Yoon, I. (2015). Elevated transmission of upper respiratory illness among new recruits in military barracks in Thailand. *Influenza and Other Respiratory Viruses*, 9(6), 308–314. <https://doi.org/10.1111/irv.12345>
- Moeller, J., Brackett, M. A., Ivcevic, Z., & White, A. E. (2020). High school students’ feelings: Discoveries from a large national survey and an experience sampling study. *Learning and Instruction*, 66, 101301. <https://doi.org/10.1016/j.learninstruc.2019.101301>
- Riessman, C. K. (2008). *Narrative methods for the human sciences* (pp. x, 251). Sage Publications, Inc.
- Ryan, M., & Ryan, M. (2013). Theorising a model for teaching and assessing reflective learning in higher education. *Higher Education Research & Development*, 32(2), 244–257. <https://doi.org/10.1080/07294360.2012.661704>
- Sousa, R. C. R. D., Paula, W. K. A. S. D., Alves, F. A. P., Albuquerque, M. I. N. D., Albuquerque, G. A., & Coriolano-Marinus, M. W. D. L. (2023). Continuing education on child development in primary care: Healthcare workers’ perspectives. *Revista Da Escola de Enfermagem Da USP*, 57, e20230189. <https://doi.org/10.1590/1980-220x-reeusp-2023-0189en>
- Sulaiman, A. (2022). *Remaja: Antara Kemandirian dan Kenakalan*. Penerbit Bina Ilmu.

Weiner, B. (1986). *An Attributional Theory of Motivation and Emotion*. Springer US.  
<https://doi.org/10.1007/978-1-4612-4948-1>.

### **Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

# The Influence of Demonstration Lessons on Teachers' Conceptions of Good Teaching

Hua Fang<sup>1</sup>

<sup>1</sup> Hangzhou Normal University, Hangzhou, China

Correspondence: Hua Fang, Hangzhou Normal University, Hangzhou, China.

doi:10.63593/RAE.2788-7057.2026.06.006

## Abstract

Demonstration lessons are widely used in teacher development because they make teaching practices visible and provide teachers with concrete examples of classroom organization, interaction, and instructional design. Yet the lessons selected for observation are often highly prepared and presented under conditions that differ from everyday teaching. This paper examines how repeated exposure to such lessons may influence teachers' conceptions of good teaching. It focuses on the preference for visible classroom performance, the smoothness produced by highly structured interaction, and the reduced presence of uncertainty in demonstration settings. The discussion also considers how these features may be transferred into ordinary classroom practice, especially when teachers begin to associate good teaching with complete procedures, active participation, and immediately observable outcomes. Drawing on policy initiatives and typical classroom situations, the paper argues that the main issue is not demonstration lessons themselves, but the way they can become implicit standards for judging teaching quality. When external forms are given too much weight, teachers may pay less attention to contextual judgment, student differences, and the unfinished or unpredictable nature of real classroom learning. Demonstration lessons are most useful when treated as materials for professional reflection rather than as models to be reproduced.

**Keywords:** demonstration lessons, good teaching, teacher professional judgment, classroom performance, teacher development, pedagogical practice

## 1. Introduction

Demonstration lessons occupy a familiar place in teacher development. They are used in teaching competitions, professional training, school-based research activities, and the sharing of exemplary practice. For many teachers, especially those in the early stages of their careers, observing a well-designed lesson is one of the most direct ways to understand how teaching can be organized. A demonstration lesson shows not only what a teacher does, but also what a successful classroom is expected to look like.

This practical value should not be underestimated. Compared with abstract descriptions of pedagogy, a visible classroom gives teachers something concrete to observe: the sequence of activities, the wording of questions, the organization of interaction, the use of teaching materials, and the way a lesson is brought to a close. Problems begin to appear, however, when a demonstration lesson is treated less as one possible example and more as a standard of good teaching. The conditions under which such lessons are prepared are often quite different from those of ordinary classroom teaching. They may involve repeated lesson polishing, peer feedback, careful timing, selected materials, and a strong awareness of external observation. The final lesson is usually expected to be coherent, engaging, and easy to evaluate.

These features can gradually affect how teachers understand teaching quality. A classroom with active participation, smooth transitions, clear stages, and visible learning outcomes is more easily recognized as

successful. By contrast, hesitation, incomplete answers, changes of plan, or slower learning processes are less likely to fit the conventional image of an exemplary lesson. Yet such uncertainty is common in everyday teaching. Students do not always respond as expected, and teachers often need to change direction in response to what actually happens in the classroom.

The issue, then, is not whether demonstration lessons are useful. They clearly have value as resources for observation and professional discussion. What deserves closer attention is the influence they may have on teachers' conceptions of good teaching. If repeated exposure to highly polished lessons encourages teachers to equate good teaching with visible performance, the judgment of teaching quality may gradually move away from questions of student understanding, classroom context, and professional responsiveness.

This paper examines that influence. It considers how demonstration lessons help produce particular images of good teaching, why observable classroom features often receive greater attention, and how these standards may be carried into everyday practice. Particular emphasis is placed on the possible weakening of context-based professional judgment when teachers begin to reproduce the outward form of exemplary lessons. The discussion does not reject demonstration teaching as a professional practice. Rather, it argues for a more careful distinction between a lesson that is effective for demonstration and teaching that is genuinely appropriate for a particular group of students.

## **2. Demonstration Lessons and the Making of "Good Teaching"**

Teachers do not develop an understanding of good teaching only from educational theory or formal training. Much of that understanding comes from seeing how other teachers teach. A lesson that is publicly presented as exemplary carries a particular authority: it is not simply another classroom to observe, but a classroom that has already been recognized as worth learning from. For beginning teachers, this kind of example can be especially influential because ideas such as student-centered teaching, effective questioning, or meaningful interaction are easier to grasp when they are attached to a concrete lesson.

This helps explain why demonstration lessons have remained important in teacher development. Their value lies partly in visibility. Teachers can see how a lesson begins, how questions are sequenced, how activities are connected, and how the teacher deals with students' responses. These are details that general pedagogical principles cannot always convey. Observation also gives teachers a language for discussing practice with colleagues. A teaching method that appears abstract in a training document may become much clearer once it is seen in a classroom.

The national initiative known as *One Teacher, One Excellent Lesson; One Lesson, One Excellent Teacher* offers a useful example. Launched by the Ministry of Education in 2014, the program encouraged teachers to develop and share high-quality lessons and sought to build a large body of exemplary classroom resources. The initiative later accumulated millions of uploaded lessons, with selected lessons identified at provincial and national levels and made available for teachers to consult. Its significance was not limited to resource sharing. By selecting, displaying, and circulating certain lessons as "excellent," the program also provided teachers with visible examples of what successful teaching could look like.

The evaluation of such lessons matters here. In the early stage of the initiative, official evaluation indicators were issued for the selection of "excellent lessons." Once a lesson enters a process of recommendation and selection, teaching is no longer seen only by the students who happen to be in that classroom. It is also viewed by peers, teaching researchers, and evaluators. Features that can be recognized within a limited period of observation naturally become more noticeable: a clear structure, purposeful activities, effective use of resources, student participation, and an identifiable learning outcome. None of these is undesirable. Most are reasonable features of competent teaching. The difficulty begins when they are taken together as a relatively fixed picture of what a good lesson should look like.

Repeated exposure to exemplary lessons can make that picture increasingly familiar. Teachers may come to expect a good lesson to have a strong opening, carefully connected stages, frequent interaction, and a visible conclusion showing what students have learned. A lesson that looks less complete may then appear weaker even when the teacher has made sound decisions in response to students' actual difficulties. The influence is subtle. Teachers are rarely told directly that every successful lesson must follow the same pattern. The standard is learned through observation, comparison, discussion, and the repeated circulation of particular forms of successful practice.

This is why demonstration lessons are more than neutral teaching resources. They participate in the formation of professional expectations. What teachers repeatedly see being praised can gradually enter their own judgment of teaching quality. The issue is not that exemplary lessons should cease to exist, nor that teachers should avoid learning from them. Their professional value is clear. What needs closer examination is the kind of image of good teaching that becomes most visible through them, and what may receive less attention when teaching is

viewed primarily through exemplary performance. That question becomes more important once the focus moves from the existence of demonstration lessons to the particular qualities that are easiest to display and evaluate.

### **3. From Good Teaching to Visible Teaching**

#### *3.1 The Preference for Observable Classroom Performance*

A demonstration lesson has to be seen before it can be judged, and this condition affects what counts as evidence of good teaching. Observers can see how the teacher organizes activities, whether students respond, how smoothly the lesson develops, and what appears to have been achieved by the end. What they cannot easily see is the longer learning process behind the lesson: persistent misunderstandings, slow progress, or decisions that only make sense in relation to a particular group of students. Demonstration settings therefore tend to give greater weight to qualities that can be recognized within a limited period of observation.

The national “Basic Education Excellent Courses” initiative provides a useful example. The program is intended to collect and circulate high-quality teaching resources. In the 2024 selection requirements, submitted courses were expected to demonstrate clear objectives, a complete teaching process, sufficient resources, and classroom effectiveness, while also meeting technical standards for production. In the 2023 selection, more than 220,000 teachers from over 60,000 schools participated, with selected lessons eventually recognized at the national level. Once such lessons are circulated as exemplary resources, they also become visible references for how successful teaching is expected to appear.

The standards themselves are not unreasonable. A lesson with unclear objectives or weak instructional logic would hardly serve as a useful professional model. The more difficult issue is that some aspects of teaching are much easier to display than others. A clear sequence can be observed immediately. Student participation can be heard, an activity can be seen to produce a result, and a well-prepared resource leaves a visible impression. A teacher who spends additional time on a difficult idea because students remain confused may be making a sound pedagogical decision, but that decision is less easy to present as evidence of a complete and successful lesson.

Classroom participation illustrates this difference. A lesson in which students answer frequently, groups report quickly, and activities move continuously can appear highly engaged. Another class may contain fewer visible interactions because students are reading, thinking, revising an answer, or struggling with a difficult concept. The second classroom may look less impressive to an observer even though the learning involved is no less serious. The same is true of lesson structure. Everyday teaching often requires teachers to slow down, abandon an activity, or return to an earlier point when students do not understand as expected.

This difference matters because demonstration lessons can gradually influence the professional image of good teaching. What is selected and repeatedly praised becomes familiar. Clear procedures, active interaction, visible outcomes, and smooth progression may begin to function as informal indicators of teaching quality even when no formal rule requires them. Teachers can then come to associate the outward form of an exemplary lesson with good teaching itself.

The distinction, then, is between what is valuable in teaching and what is easy to display as valuable. These two categories often overlap, but they are not identical. Once visible performance becomes a major reference for judging classroom quality, the next question is whether interaction itself begins to be shaped by the need to appear successful and well controlled.

#### *3.2 Scripted Interaction and the Appearance of Smooth Teaching*

Smooth classroom interaction is another feature that often stands out in demonstration lessons. Questions are answered quickly, activities move forward without much interruption, and students seem to understand what is expected at each stage. This can create a strong impression of teaching competence. Yet ordinary classrooms are rarely this predictable.

Part of this difference comes from preparation. Demonstration lessons are often revised repeatedly before they are presented. Questions may be adjusted, transitions improved, and possible student responses anticipated in advance. The 2025 “Basic Education Excellent Courses” selection process, for example, allowed county-level organizers to arrange trial teaching before final recording. Such preparation is professionally useful, but it can also reduce the uncertainty that normally exists in classroom interaction.

Student responses may become more predictable even when they are not literally rehearsed. Teachers who have already tested a lesson usually know where students may hesitate and can redesign questions to produce clearer answers. Activities can also be arranged so that a particular result is more likely to appear within the available time. The lesson remains interactive, but the interaction is more controlled than it would be in everyday teaching.

The difficulty is that uncertainty often provides important information about learning. A wrong answer may reveal a misconception. Silence may show that students need more time. An unexpected response may force the teacher to reconsider the planned direction. These moments are less convenient in a lesson expected to proceed

smoothly, but they are often where professional judgment becomes necessary.

Smoothness, then, needs to be viewed with some caution. A highly controlled lesson may look effective because teacher action and student response fit together neatly, while a less polished lesson may involve more genuine adjustment to students' understanding. The problem appears when this polished form begins to stand in for successful teaching itself. Once that happens, the ordinary uncertainty of classroom learning can start to look like a weakness rather than a normal part of teaching.

### *3.3 The Marginalization of Uncertainty in Classroom Teaching*

Uncertainty is difficult to avoid in ordinary classroom teaching. Students may misunderstand a question, take longer than expected to complete a task, or respond in a way that the teacher did not anticipate. A discussion may fail to develop, while an apparently simple point can suddenly require much more explanation. These situations make teaching less orderly, but they also give teachers information about how students are actually learning. A classroom in which everything happens exactly as planned is not necessarily a classroom in which learning has been understood well.

This point is easily overlooked when demonstration lessons become an important reference for good teaching. Lessons prepared for observation usually have limited time and need to make their instructional value recognizable. Unexpected events can interfere with that purpose. A long silence after a question, several incorrect answers, or an unfinished task may be pedagogically meaningful, yet these moments are harder to accommodate in a lesson expected to appear complete. Uncertainty is gradually treated as something to be reduced through better planning rather than something that teachers sometimes need to work with.

Current curriculum expectations offer a useful contrast. The *Compulsory Education Curriculum Plan (2022 Edition)* emphasizes subject practice, learner-centered environments, differentiated teaching, and the process through which students discover problems, solve them, construct knowledge, and apply what they have learned. It also stresses connections with students' experience and real-life situations. These ideas leave room for a learning process that is not entirely predictable. If students are genuinely exploring a problem, their responses cannot always be known in advance. Differences among students also mean that the same activity will not necessarily produce the same result at the same speed.

The tension is quite clear. A demonstration lesson often benefits from a visible sense of progress, whereas inquiry and differentiated learning may produce uneven progress. A teacher may need to stay with a student's unexpected answer, change the planned sequence, or leave part of a question unresolved until the next lesson. Judged against a polished demonstration, such decisions can look inefficient. Judged against the needs of the students in front of the teacher, they may be entirely appropriate.

What tends to disappear from an overly controlled image of good teaching is precisely this kind of professional response. Teachers do not only carry out lesson plans; they read the classroom while teaching. They decide whether a silence means confusion or serious thought, whether an error should be corrected immediately, and whether completing the planned activity matters more than dealing with a difficulty that has just emerged. Much of this work is difficult to display because its value depends on context.

The problem is not that demonstration lessons should deliberately become disorganized. Careful preparation remains part of good teaching. What needs more attention is the assumption that uncertainty itself signals weak teaching. When polished completion becomes too closely associated with quality, the less orderly parts of classroom learning can be pushed to the margins. Yet these are often the moments in which teachers have to exercise the professional judgment that a demonstration lesson is supposed to help develop.

## **4. The Influence on Teachers' Everyday Practice and Professional Judgment**

### *4.1 The Transfer of Demonstration-Lesson Standards to Everyday Teaching*

The influence of demonstration lessons becomes more visible when teachers begin to carry their standards back into ordinary classrooms. This transfer is understandable. Demonstration lessons are often presented as examples of successful practice, so teachers naturally pay attention to the features that appear to make them successful. Clear lesson stages, frequent interaction, well-timed activities, and visible outcomes can gradually become reference points for planning daily teaching.

The problem is that the conditions of everyday teaching are different. A demonstration lesson is usually prepared around a particular class, a fixed period of time, and a clear presentation goal. Daily teaching is less controlled. Students vary in prior knowledge, attention, confidence, and pace of learning. A structure that works well in a demonstration setting may not produce the same result when it is transferred directly to another classroom.

A common situation can be seen among less experienced teachers. After observing an award-winning or widely shared demonstration lesson, a teacher may try to reproduce its structure in a regular class. The lesson may include several short activities, frequent questioning, group discussion, and a final presentation of learning

outcomes. On paper, the design looks active and student-centered. In practice, however, students may need more time to think, group work may not lead to an immediate answer, and some activities may contribute little to the actual learning goal. If the teacher still tries to complete every planned stage, the lesson can become busy without becoming more effective.

This kind of transfer is not simple imitation. It reflects a deeper shift in what the teacher expects a good lesson to look like. Activities that are highly visible in demonstration lessons can begin to feel necessary even when the content does not require them. Teachers may worry that a lesson with fewer activities, longer explanation, or extended individual thinking appears too traditional or insufficiently engaging. The standard of quality then moves from the question of what students need toward the question of whether the lesson resembles an accepted model.

There is also a tendency to preserve lesson completeness. A teacher may continue with the next task because the lesson plan requires it, even when student responses suggest that an earlier point has not been fully understood. The decision is easy to explain. In demonstration lessons, the ability to move through the planned sequence is often part of the visible impression of competence. Once this expectation enters daily teaching, adjustment can feel like failure rather than professional responsiveness.

The transfer of standards is especially important for teachers who are still developing their own professional judgment. Exemplary lessons provide useful models, but the visible form of a lesson is easier to copy than the reasoning behind it. A teacher can reproduce a questioning sequence or activity pattern without knowing why it worked with a particular group of students. What appears to be methodological learning may therefore remain at the level of form.

The issue is not whether teachers should learn from demonstration lessons. They should. The more difficult question is what exactly should be transferred. A useful lesson does not offer a fixed structure for repetition. Its greater value lies in the decisions behind the structure: why the teacher chose one activity rather than another, why a question was asked at a particular moment, and how the lesson responded to the needs of that class. When these decisions are overlooked, the outward form of good teaching can travel more easily than the professional judgment that produced it.

#### *4.2 The Pressure to Reproduce an Ideal Classroom*

Once demonstration lessons become a familiar image of successful teaching, teachers may begin to compare ordinary classrooms with that image. The comparison is rarely fair. A demonstration lesson is usually prepared for observation, while everyday teaching takes place under routine conditions and must deal with uneven participation, incomplete understanding, time pressure, and unexpected interruptions. Yet the polished lesson can still become the standard against which ordinary teaching is judged.

This creates a subtle form of pressure. Teachers may feel that a successful class should be active, orderly, and visibly productive from beginning to end. When students remain silent, an activity takes too long, or a discussion fails to produce the expected answer, the lesson may seem unsuccessful even when the teacher is responding appropriately to the situation. Normal classroom difficulty is then interpreted as a gap between actual teaching and an ideal model.

The gap can be particularly frustrating because it is difficult to remove through preparation alone. Teachers can plan questions and activities carefully, but they cannot fully predict how students will respond on a particular day. A lesson may work well with one class and poorly with another. Differences in prior knowledge, motivation, classroom relationships, and learning pace all affect what happens. Treating a demonstration lesson as a reproducible model tends to hide these differences.

This pressure may also change teachers' priorities during a lesson. Instead of asking whether students need more time, a teacher may worry that the class is moving too slowly. Instead of allowing a difficult discussion to remain open, the teacher may bring it to a quick conclusion so that the planned sequence can continue. The lesson becomes easier to manage as a performance, but less responsive to what students are showing.

The more serious issue is how teachers come to judge their own professional competence. If ordinary classroom uncertainty is repeatedly compared with highly prepared exemplary lessons, teachers may undervalue teaching that is quiet, incomplete, or less visually impressive. A lesson in which the teacher notices confusion and changes direction may look less polished, yet such adjustment can require more professional skill than following a plan without interruption.

Demonstration lessons become problematic when they create an ideal that everyday teaching is expected to reproduce. Their value lies in offering possibilities for reflection, not in setting a fixed image of classroom success. The pressure to imitate that image can narrow teachers' understanding of what effective teaching may look like and prepare the ground for a deeper problem: professional judgment becomes increasingly tied to

external forms rather than to the needs of a particular classroom.

#### *4.3 The Weakening of Context-Based Professional Judgment*

The deeper influence of demonstration lessons appears when teachers begin to rely on external models more than on the conditions of their own classrooms. Professional judgment depends on context. The same activity may work well with one group of students and fail with another; a question that produces productive discussion in one lesson may create confusion in a different class. Teaching therefore requires continuous decisions about pace, explanation, participation, and the amount of support students need. These decisions cannot be settled in advance by a model lesson.

Demonstration lessons can still contribute to this judgment, but only when teachers look beyond the visible structure. Problems arise when the form of an exemplary lesson is treated as evidence of quality in itself. A teacher may begin to assume that good teaching should contain frequent interaction, varied activities, group work, or a clear final product because these features appear repeatedly in successful examples. The question of whether they are necessary for a particular lesson can receive less attention.

This is especially significant for teachers who have not yet developed a stable sense of what works in their own classrooms. Copying a visible technique is much easier than understanding the reasoning behind it. A teacher can reproduce a sequence of questions, a group task, or a digital activity without knowing why the original teacher used it at that moment and with those students. What looks like professional learning may then remain close to methodological imitation.

Context-based judgment works differently. It requires the teacher to notice what students are doing and to accept that the original plan may need to change. A slower class may need fewer activities and more explanation. A group that already understands a concept may need a more demanding task rather than additional practice. Some students may benefit from discussion, while others need time to work through an idea individually. None of these choices is particularly impressive as a visible teaching technique. Their value comes from their fit with the situation.

The danger of an idealized model of good teaching is that it can make such adaptation appear secondary. Teachers may become more concerned with whether a lesson looks complete than with whether the planned sequence still makes sense. A decision to drop an activity, extend an explanation, or allow students more time can feel like a failure to deliver the intended lesson. Yet these are often precisely the decisions that show professional awareness.

Professional judgment is not the rejection of planning or established methods. It is the ability to use them selectively. Teachers need models, examples, and shared standards, but they also need enough confidence to depart from them when classroom evidence demands it. Good teaching cannot be reduced to a recognizable arrangement of activities because its quality depends partly on how well the teacher reads a specific group of students.

The concern, then, is not that demonstration lessons directly weaken professional judgment. The effect is more indirect. When exemplary forms are repeatedly treated as signs of quality, teachers may gradually give more authority to the model than to the classroom in front of them. That shift matters because professional growth depends less on reproducing successful lessons than on learning when a method fits, when it does not, and what should be changed when reality refuses to follow the plan.

#### *4.4 Rethinking the Meaning of Classroom Success*

The discussion above also raises a basic question: what should count as a successful lesson? In demonstration settings, success is often associated with visible completion. Students participate, activities run as planned, and the lesson ends with a clear result. These features are easy to recognize and evaluate. They are useful, but they do not capture the whole quality of classroom teaching.

A successful lesson may sometimes look less complete. Students may still be uncertain at the end of a discussion, or the teacher may decide to spend more time on one difficulty and leave part of the original plan unfinished. Such a lesson can appear weaker when judged by pace and structure, yet it may respond more closely to what students actually need. Learning does not always produce an immediate and observable result within a single class period.

This matters for teachers' own judgments of their practice. If classroom success is defined mainly by smoothness, participation, and completion, teachers may overlook quieter signs of learning. A student who revises an earlier misunderstanding, a class that begins asking better questions, or a teacher who recognizes that the planned task is not working may represent meaningful progress even when the lesson is not polished.

The idea of success therefore needs to leave room for difficulty, adjustment, and partial understanding. These are not signs that teaching has failed. They are often part of the process through which teachers discover what

students know and what still needs attention. A classroom that looks imperfect can still be educationally productive.

Demonstration lessons are useful when they broaden teachers' understanding of possible practice. They become less helpful when they narrow success into a fixed classroom appearance. Good teaching is better judged by the quality of decisions made in relation to students and content than by how closely a lesson resembles an ideal model.

### 5. Rethinking Demonstration Lessons and the Meaning of Good Teaching

The value of demonstration lessons should not be judged only by whether the lesson itself appears successful. Their more useful role lies in what they allow teachers to examine. A well-prepared lesson can make instructional decisions visible and provide a starting point for professional discussion. The difficulty begins when observation stops at the level of form. Teachers may remember the sequence of activities, the questioning pattern, or the use of technology, while the reasoning behind those choices receives much less attention.

A more productive use of demonstration lessons would treat them as cases for interpretation rather than models for replication. Observers need to ask why a particular activity was chosen, what assumptions the teacher made about the students, and how the lesson might have changed if those students had responded differently. These questions shift attention away from whether a lesson looks complete and toward the professional decisions that produced it. They also make it easier to recognize that a successful practice in one classroom may need substantial adjustment in another.

This change in perspective also affects the meaning of good teaching. Good teaching cannot be identified through one visible classroom pattern. Some lessons require sustained explanation; others depend on discussion or practical work. A quiet classroom may contain serious thinking, while an active classroom may produce little understanding. The number of activities, the frequency of student responses, or the smoothness of transitions can offer information about a lesson, but none of them should function as a substitute for judging what students are actually learning.

Demonstration lessons could also make greater room for the less polished side of teaching. Discussion after a lesson, for example, does not need to focus only on successful techniques. Teachers may learn just as much from a question that failed, an activity that took too long, or a moment when the teacher changed the original plan. Such episodes reveal the uncertainty of teaching and show how professional judgment works under real conditions. Removing all signs of difficulty may make a lesson easier to admire, but it can also make professional practice look more predictable than it really is.

The idea of an exemplary lesson is still useful, but "exemplary" need not mean that others should reproduce its visible structure. A demonstration lesson can be exemplary because it shows careful reasoning, sensitivity to students, and a willingness to adjust when the situation changes. Seen in this way, its purpose is not to define a single image of good teaching. It is to give teachers material from which they can develop their own judgment.

Teachers ultimately need more than models of successful classroom performance. They need the capacity to decide what is appropriate for particular students, content, and circumstances. Demonstration lessons can support that development when they open up questions rather than close them down. The strongest lesson to take from an exemplary classroom may not be how to make another lesson look similar, but how to think more carefully about what good teaching requires in a different one.

### References

- Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, 18(8), 947–967. doi:10.1016/S0742-051X(02)00053-7.
- Fenstermacher, G. D., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record*, 107(1), 186–213. doi:10.1111/j.1467-9620.2005.00462.x.
- Grierson, A. L., & Gallagher, T. L. (2009). Seeing is believing: Creating a catalyst for teacher change through a demonstration classroom professional development initiative. *Professional Development in Education*, 35(4), 567–584. doi:10.1080/19415250902930726.
- Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, 86(4), 945–980. doi:10.3102/0034654315626800.
- Liu, H., Zhang, Q., & Bai, B. (2019). Are public lessons in China teachers' shows? An investigation of English teachers' beliefs. *Journal of Education for Teaching*, 45(5), 608–610. doi:10.1080/02607476.2019.1675357.
- Ministry of Education of the People's Republic of China. (2014). Notice on carrying out the 2014 "One Teacher, One Excellent Lesson; One Lesson, One Excellent Teacher" initiative.
- Ministry of Education of the People's Republic of China. (2022). Compulsory education curriculum plan (2022

- edition). Beijing Normal University Publishing Group.
- Ministry of Education of the People's Republic of China. (2024). Notice on the selection of 2024 Basic Education Excellent Courses.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307–332. doi:10.3102/00346543062003307.
- Rhodes, G., Hallman-Thrasher, A., & Schultz, K. T. (2019). Mathematics teacher developers' views of a demonstration class. *The Mathematics Educator*, 28(1), 61–89. doi:10.63301/tme.v28i1.2047.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23. doi:10.17763/haer.57.1.j463w79r56455411.
- Wu, H., & Clarke, A. (2018). The Chinese “Open Class”: A conceptual rendering and historical account. *Asia Pacific Journal of Education*, 38(2), 214–228. doi:10.1080/02188791.2018.1460260.
- Wu, J., & Shen, P. (2024). Facilitating teachers' professional learning in open class: A qualitative case study in China. *Humanities and Social Sciences Communications*, 11, Article 1752. doi:10.1057/s41599-024-04348-9.

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).