

Language Teachers' Efficacy in Using Generative Language AI in Instruction: A Mixed-Method Explanatory Sequential Study

Therese Marie Bedonia¹, Marijo D. Chua²

¹Department of Education Division of Negros Occidental Philippines

²Carlos Hilado Memorial State University Negros Occidental, Philippines

ABSTRACT	ARTICLE DETAILS
Artificial Intelligence (AI), particularly generative language AI, is rapidly transforming education. This study explored the efficacy of language teachers in integrating generative language AI into instruction within a large schools division in the Philippines during the 2024–2025 school year. Using a mixed-method explanatory sequential design, data were collected through surveys and focus group discussions. Quantitative data were analyzed using mean, standard deviation, and the Mann-Whitney U test, while qualitative data underwent thematic analysis using Lichtman's 3C's model. Findings showed that language teachers demonstrated high efficacy in using generative AI, particularly in four domains: adoption, application, adaptation, and ethical consideration. Notably, there were no significant differences in efficacy based on demographic factors such as age, curricular category, educational attainment, sex, and teaching experience. The qualitative phase revealed four themes: AI as a catalyst for teaching efficiency, a driver of teacher productivity and work-life balance, a means of democratizing access to educational resources, and a double-edged sword in writing instruction. The meta-inference demonstrated alignment between the quantitative and qualitative results, highlighting teachers' readiness and optimism toward AI integration. However, challenges such as student over-dependence on AI and limited internet connectivity remain significant obstacles. To support effective AI adoption, stakeholders must improve access to technology and implement strategies to mitigate negative impacts.	Published On: 28 July 2026
KEYWORDS: Artificial Intelligence, Efficacy, Instruction, Language Teachers	Available on: https://ijiissh.com/

INTRODUCTION

Artificial Intelligence (AI) is increasingly becoming a transformative force across various sectors, influencing industries, reshaping daily activities, and redefining societal norms (Trabelsi, 2024). Among the most impactful forms of AI is Generative Language AI, which produces original, text-based content—such as articles, scripts, and dialogues—in response to user prompts (Marteanu, 2023). According to Malec (2024), Generative Language AI enhances productivity, creativity, and efficiency across multiple domains. In education, UNESCO (2023) underscores its potential to personalize learning experiences, streamline content creation, enhance student engagement, provide real-time feedback, and reduce administrative burdens. Teachers increasingly recognize AI as a supportive tool, particularly in tracking student progress and tailoring instruction to individual needs (Ugyun, 2024).

A Needs Assessment Survey conducted by the Human Resource Development Section (2024) of a large school division Office in the Philippines for the Mid-Year In-Service Training revealed a critical challenge: teachers struggle to effectively navigate and integrate rapidly evolving AI tools into their instructional practices. This issue is particularly evident among language teachers handling writing-intensive subjects like Reading and Writing Skills, Practical Research, and Creative Writing, where students frequently rely on AI-generated content for assignments and performance tasks.

Despite the growing body of international literature highlighting the potential of Generative Language AI in enhancing teaching efficiency, personalizing instruction, and reducing administrative workload, there remains a notable lack of empirical research exploring how language teachers, particularly within the Philippine public school system, utilize these tools in actual classroom settings. Furthermore, existing studies have yet to investigate how efficacy in AI integration varies across teacher demographic profiles, and how these efficacy levels are experienced and explained in practice.

Language Teachers' Efficacy in Using Generative Language AI in Instruction: A Mixed-Method Explanatory Sequential Study

This study sought to fill this gap by employing a mixed-methods explanatory sequential design, beginning with the collection of quantitative data on the efficacy of language teachers in utilizing generative language AI in instruction. Subsequently, qualitative data was gathered to explore the underlying reasons for the observed efficacy levels. By combining these data sources, the study aimed to provide comprehensive insights into teachers' experiences, challenges, and successes in integrating AI into their teaching practices, ultimately enhancing AI integration in language education.

MATERIALS AND METHODS

The study employed the mixed-method explanatory sequential research design which involved two distinct phases: an initial quantitative phase followed by a qualitative phase. In this approach, researchers first collect and analyze quantitative data to identify general trends or patterns (Toyon, 2021). Subsequently, qualitative data is gathered to provide deeper insights and explanations for the quantitative findings. This process allows for a comprehensive understanding of the research problem by integrating numerical data with contextual narratives.

From a practical standpoint, the quantitative phase allowed for the collection of measurable data on key variables such as age, curricular category, educational attainment, sex, and teaching experience. This provided an objective basis for identifying trends and patterns in AI utilization. However, understanding the practical challenges and motivations behind these trends required qualitative exploration. The qualitative phase, conducted through focused group discussions, helped uncover specific barriers, facilitators, and contextual factors influencing AI adoption in instructional settings.

At the procedural level, the explanatory sequential design was strategically structured to enhance the study's depth and validity. By first analyzing quantitative data, the study identified key trends and gaps that required further exploration. The qualitative phase was then designed to explain these findings, ensuring that statistical results were contextualized and interpreted meaningfully. This step-by-step approach strengthened the study's credibility by providing both empirical evidence and rich, narrative-driven insights, leading to a well-rounded and substantiated conclusion.

Meanwhile, for the quantitative phase of the study, the respondents consisted of English language teachers from a large school division in the Philippines. The total population comprised 711 teachers in the Junior and Senior High Schools dispersed in the 34 districts comprising the schools division. Using the G-Power formula, a statistically representative sample size of 159 was selected. To ensure the representative of each district, a proportional stratified sampling technique was used. Proportional stratified sampling is a technique where a population is divided into strata (distinct subgroups) based on specific characteristics, and samples are drawn from each stratum in proportion to their representation of the overall population, which ensures that each subgroup is adequately represented, leading to more precise and reliable estimates (Reddy & Khan, 2023). In addition, respondents were selected based on their eligibility on the following criteria: they must be English language teachers in either junior or senior high school within the Department of Education (DepEd) with at least one year of teaching experience. Teachers with at least one year of service have had sufficient exposure to classroom realities, teaching responsibilities, and professional development activities aligned with the PPST. This makes them more capable of reflecting on their efficacy in using tools like Generative Language AI in actual instruction. These teachers have likely undergone at least one full school year of curriculum implementation, student assessment, and classroom management. Thus, their insights into integrating AI tools are based on sustained teaching experience rather than theoretical or limited exposure. Eligible respondents must also hold a permanent, regular position in the department and possess degrees related to English language teaching. Several key variables had been considered to understand the efficacy of language teachers' utilization of generative language AI in instruction such as: age, curricular category, educational attainment, sex, and teaching experience. These provided insights into the diverse demographic and professional backgrounds of the respondents, which could influence their ability to integrate AI tools into their teaching.

The selection of participants for the qualitative phase was guided by the results of the quantitative phase, employing a maximum variation sampling strategy to capture a wide range of perspectives. Maximum variation sampling is a type of purposive sampling technique commonly used in qualitative studies to capture a wide range of perspectives related to the phenomenon of interest (Palinkas et al., 2015). By intentionally selecting participants who represent both ends of the efficacy spectrum—those with the highest and lowest levels of proficiency—the study captured the broadest range of insights, challenges, and strategies related to AI integration. The inclusion criteria mirrored those used in the quantitative phase; however, participants were specifically selected from among the original respondents based on their efficacy levels in utilizing generative language AI in instruction. To ensure a balanced representation of perspectives, three language teachers with high efficacy and three with low efficacy were invited to participate in the Focus Group Discussion (FGD).

An analysis of district-level performance across the evaluation areas—Adoption, Application, Adaptation, Ethical Consideration, and Efficacy in Utilization—revealed significant differences. District 9 consistently recorded the highest scores, while District 32 registered the lowest across several domains. Consequently, the FGD was conducted in these two districts to capture contrasting insights and experiences based on maximum variation in performance.

To facilitate the identification and recruitment of suitable participants, assistance from key gatekeepers was sought. Their insights and familiarity with the teachers' performance and instructional practices were instrumental in selecting participants who met the criteria and provided valuable contributions to the study.

The quantitative data in the study was collected through a researcher-made questionnaire which consisted of two (2) parts. The first part focused on gathering the demographic profile of the respondents including their age, curricular category, educational attainment, sex, and teaching experience. This demographic profile helped contextualize the findings and identify potential patterns and differences across respondent groups. The second part of the questionnaire assessed the level of utilization of generative language AI in instruction through a Likert scale. It typically consists of a series of statements to which respondents indicate their level of agreement or disagreement on a symmetric agree-disagree scale, often with five or seven points. This format allows for the quantification of subjective perceptions, facilitating statistical analysis. This section presents thirty-two (32) statements derived from completed research studies closely related to the present investigation. These studies include works by Ghimire and Edwards (2024), Petterson et al. (2024), The Open Innovation Team and Department for Education (2024), Al-Abdullatif and Mohammed (2024), Fitria (2024), Su and Li (2021), Ruiz-Roxas et al. (2024), Lee et al. (2022), Kaplan-Rakowski (2023), and Batool et al. (2024). The selected statements provide relevant insights and empirical support that informed the development of the research instrument. The statements were distributed across four key indicators: adoption, application, adaptation, and ethical considerations. Each indicator was assessed using eight (8) statements. Specifically, Items 1–8 focused on adoption, Items 9–16 on application, Items 17–24 on adaptation, and Items 25–32 on ethical considerations. Respondents rated each statement using a 5-point Likert scale, ranging from strongly agree (5) to strongly disagree (1).

Before data collection, the researcher-made questionnaire underwent rigorous validity and reliability testing to ensure accuracy and consistency. High validity indicates that the results accurately reflect real properties, characteristics, and variations within the physical or social world (Kamper, 2019). For content validity, the questionnaire was evaluated by a panel of fifteen (15) experts, consisting of language educators and IT professionals with expertise in generative AI applications in education. Lawshe's Content Validity Ratio (CVR) was applied to assess how well each item represents the construct of language teachers' efficacy in utilizing generative language AI in instruction (Jeldres et al., 2023). The Content Validity Ratio (CVR) for each item was then calculated and compared to Lawshe's critical values, which depend on the number of panelists involved. For a panel size of 15, the critical CVR value is 0.60. Based on the panel members' ratings, the Content Validity Index (CVI) was calculated at 0.979 which confirms that the questionnaire items are well-validated and do not require significant revision or removal. In addition to the quantitative CVR analysis, qualitative feedback from the experts was considered to refine the wording, clarity, and overall comprehensiveness of the questionnaire. This iterative process ensured that the instrument was both valid and aligned with the study's objectives. Following the completion of content validity procedures and the incorporation of expert feedback, the questionnaire underwent reliability testing. For the reliability testing, the researcher-made questionnaire underwent internal consistency testing, which assesses how well the items within the questionnaire are interrelated, ensuring they consistently measure the same underlying construct (McCrae et al., 2011). The questionnaire was piloted among 65 language teachers from the 34 districts in the large schools division. The cooperation of key gatekeepers—Public Schools District Supervisors, District Human Resource Development Coordinators, and School Heads—was sought to identify two (2) language teachers per district to participate in the reliability testing. It is important to note that these teachers were excluded from the final data collection to prevent any bias. Once the questionnaires were collected, Cronbach's Alpha was calculated for each section to assess the internal consistency of the items. A Cronbach's Alpha value of 0.70 or higher is considered acceptable, indicating good reliability. Upon analysis, the questionnaire achieved a Cronbach's Alpha of 0.976, signifying excellent reliability and strong internal consistency among the items. Through validity and reliability testing, the questionnaire was proven to be both accurate and consistent in measuring language teachers' efficacy in using generative language AI in instruction.

Meanwhile, the qualitative data for the study was collected through Focused Group Discussion (FGD). A Focus Group Discussion (FGD) is a qualitative research method where a small group of participants engages in a guided discussion on a specific topic to generate data through their interactions (Bachtar et al., 2024). The researcher served as the primary instrument in facilitating the discussion, supported by tools such as an audio recorder to ensure accuracy and completeness. By guiding the conversation with a clear focus on the research objectives while fostering a safe and open environment, participants were encouraged to express their thoughts freely. To capture the full scope of the discussion, an audio recorder was used to document participants' responses verbatim. Complementing this, the researcher also took field notes to record non-verbal cues, group dynamics, and contextual observations that could enrich the data. This combination of recorded dialogue and observational insights ensured the depth, authenticity, and reliability of the qualitative data collected.

The following statistical tools utilized to analyze the quantitative results, to answer the respondents' level of efficacy in utilizing generative language AI for instruction, the mean and standard deviation (SD) were employed as statistical measures, and to determine whether there are significant differences in the level of efficacy in utilizing generative language AI in instruction when

teachers are grouped according to variables such as age, curricular category, educational attainment, sex, and teaching experience, a test for normality was first conducted. Results from the Kolmogorov-Smirnov test revealed that the distribution of efficacy scores significantly deviated from normality [$D(158) = 0.135, p = 0.000$], indicating that the data was not normally distributed. Given this outcome, non-parametric statistical methods were deemed appropriate. Consequently, the Mann-Whitney U test was employed to assess significant differences in efficacy levels across the specified grouping variables. In the qualitative phase of this study, data analysis was guided by Lichtman's (2013) "3 Cs" framework—Codes, Categories, and Concepts—to ensure a systematic and rigorous approach to interpreting participants' insights.

This study outlines the rigorous methods employed throughout the qualitative phase of the study. To ensure the trustworthiness and integrity of the findings, the study adhered to established standards of credibility, dependability, confirmability, transferability, and ethical considerations (Johnson et al., 2020). By implementing these rigorous strategies, along with the ethical considerations, the qualitative phase produced reliable, credible, and transferable findings that contributed to a comprehensive understanding of language teachers' efficacy in utilizing generative language AI in instruction.

RESULTS AND DISCUSSION

The study explored two research questions through a quantitative phase followed by a qualitative phase to validate and complement the findings. The following discusses the summary of the quantitative results. The first question examined the efficacy of language teachers in using generative language AI for instruction, considering demographic profiles such as age, curricular category, educational attainment, sex, and teaching experience, across four key areas: adoption, application, adaptation, and ethical consideration. The data analysis revealed a high level of efficacy overall, with minor variations observed across demographic groups. Language teachers also reported high levels of efficacy in key areas of efficacy: adoption ($M=3.54, SD=0.71$), application ($M=3.43, SD=0.81$), adaptation ($M=3.69, SD=0.74$), and ethical considerations ($M=3.87, SD=0.84$). Teachers across both age groups—young and old—demonstrated high levels of efficacy in using generative language AI. Senior high school teachers reported slightly higher efficacy compared to their junior high school counterparts. Similarly, teachers with a master's degree exhibited marginally higher efficacy than those holding only a bachelor's degree. Male teachers also showed a slightly higher overall efficacy compared to female teachers. Interestingly, teachers with fewer years of experience reported higher efficacy levels than their more experienced peers. While these differences exist, it is important to emphasize that they are minimal, indicating a generally consistent level of competence and confidence across all demographic categories.

The second question examined whether there were significant differences in the efficacy of language teachers using AI in instruction based on their demographic profiles. The analysis revealed that no significant differences were found across all key areas of adoption, application, adaptation, and ethical consideration. This suggests that language teachers, regardless of their age, curricular category, educational attainment, sex, or teaching experience, demonstrate a comparable level of efficacy when integrating AI into their teaching practices. The lack of significant variation across these demographic factors highlights the widespread acceptance and effective use of generative language AI among teachers, regardless of individual characteristics.

For the qualitative findings, the focus group discussions illuminated four key themes regarding language teachers' experiences with generative AI: (1) AI as an amplifier of pedagogical efficiency, (2) AI as a driver of teacher productivity and work-life balance, (3) AI as a democratizer of educational resources, and (4) the two-edged sword of AI as a writing teacher. Teachers reported that AI significantly streamlines fundamental teaching functions. They highlighted AI's capacity to expedite lesson planning and automate assessment processes, with ChatGPT emerging as a particularly favored tool for its practicality in daily instruction. Furthermore, AI was seen to accelerate research for both educators and learners by facilitating easier access to and summarization of relevant literature. The study also revealed AI as a driver of teacher productivity and work-life balance. Participants indicated that AI optimizes task management across various teaching-related responsibilities. They also noted AI's support in handling often time-consuming ancillary tasks, freeing up educators to concentrate more on direct instruction. The discussions revealed AI's potential to democratize access to educational resources by providing both students and teachers with quick and easy information retrieval, particularly beneficial in settings with limited facilities. While acknowledging AI's advantages in language teaching and learning—specifically its translation functions and its ability to offer vocabulary suggestions, clarifications, and rapid information access that aids in developing students' writing skills—teachers expressed concerns about the potential for student over-dependence on these tools. Additionally, inconsistent internet connectivity was identified as a significant impediment to widespread AI adoption.

The meta-inference reveals that language teachers in this Philippine schools division demonstrate a high level of efficacy and positive experiences with generative AI, recognizing its benefits for productivity, lesson preparation, and work-life balance. This readiness to integrate AI is consistent across different teacher profiles, suggesting a collective openness to innovation. However, realizing AI's full potential hinges on addressing shared concerns about student over-reliance and inadequate internet connectivity through institutional and infrastructural support. The teachers' existing confidence and skills indicate a promising foundation for successful AI adoption with targeted interventions from educational leaders and policymakers.

CONCLUSION

This study offers valuable insights into language teachers' efficacy in utilizing generative language AI for instruction. The findings indicate a generally high level of efficacy across key areas—adoption, application, adaptation, and ethical consideration—with minimal significant variations based on demographic factors such as age, curricular category, educational attainment, sex, or teaching experience. While some minor differences were observed, language teachers' efficacy in using generative language AI is comparably high, regardless of profile. This suggests that efficacy in AI integration is not strongly tied to traditional markers of teacher experience or background. Instead, other factors, such as access to resources, training, and institutional support, maybe more influential in shaping teachers' confidence and ability to use AI effectively. This could imply that professional development should be designed to be inclusive and address the needs of all teachers, regardless of their demographic profile. Despite challenges like slow internet connections, teachers demonstrated a preference for user-friendly and accessible generative language AI tools, particularly ChatGPT. The study also underscores the importance of professional development and training to facilitate AI integration, as teachers expressed a need for more comprehensive support in adopting AI for instructional purposes. The preference for user-friendly tools highlights the importance of the Technology Acceptance Model (TAM) in the context of AI adoption. Teachers are more likely to embrace technologies that they perceive as easy to use and beneficial. This has implications for the design and selection of AI tools for educational settings. Furthermore, the expressed need for training indicates a gap between teachers' current skills and the skills required for effective AI integration. Addressing this gap through targeted professional development is crucial for successful implementation. Hence, results suggest that while demographic factors may exert some influence on specific aspects of AI adoption and usage, language teachers exhibit a consistently high level of efficacy in integrating AI into their teaching practices across all demographic groups. Consequently, targeted professional development programs aimed at deepening teachers' understanding of AI tools can further enhance their confidence and competence in leveraging AI for educational purposes. This finding has positive implications for the potential of AI in education. Since teachers across different demographics demonstrate similar levels of efficacy, it suggests that AI can be integrated into various educational contexts with a reasonable expectation of success. However, the emphasis on professional development suggests that realizing this potential requires a strategic and systematic approach.

ACKNOWLEDGMENT

This successful research's completion shows the valuable support and expertise of several key individuals throughout the research process. We sincerely thank those who offered critical insights, technical guidance, and steady encouragement. Each played a unique role in shaping the research's direction and outcomes. Their contributions, which included data analysis, methodological advice, and constructive feedback, were crucial in overcoming challenges and improving the rigor of this research.

REFERENCES

- 1) Al-Abdullatif, A. M. (2024). Modeling Teachers' Acceptance of Generative Artificial Intelligence Use in Higher Education: The Role of AI Literacy, Intelligent TPACK, and Perceived Trust. *Education Sciences*, 14(11). <https://doi.org/10.3390/educsci14111209>
- 2) Bachtiar, N. K., Fariz, M., & Arif, M. S. (2024). Conducting a Focus Group Discussion in Qualitative Research. *Innovation, Technology, and Entrepreneurship Journal*, 1(2), 94–101. <https://doi.org/10.31603/itej.11466>
- 3) Chiang, C., Jhangiani, R., & Price, P. (2015). *Research Methods in Psychology – 2nd Canadian Edition*. Pressbooks.
- 4) Gale, J., Alemdar, M., Cappelli, C., & Morris, D. (2021). A Mixed Methods Study of Self-Efficacy, the Sources of Self-Efficacy, and Teaching Experience. *Frontiers in Education*, 6. <https://doi.org/10.3389/feduc.2021.750599>
- 5) Ghimire, A., Prather, J., & Edwards, J. (2024). Generative AI in Education: A Study of Educators' Awareness, Sentiments, and Influencing Factors. <http://arxiv.org/abs/2403.15586>
- 6) Johnson, J. L., Adkins, D., & Chauvin, S. (2020). A review of the quality indicators of rigor in qualitative research. In *American Journal of Pharmaceutical Education* (Vol. 84, Issue 1, pp. 138–146). American Association of Colleges of Pharmacy. <https://doi.org/10.5688/ajpe7120>
- 7) Jones, A. (2020). Jones_ku_0099D_17560_DATA_1.
- 8) Kamper, S. J. (2019). Reliability and validity: Linking evidence to practice. In *Journal of Orthopaedic and Sports Physical Therapy* (Vol. 49, Issue 4, pp. 286–287). Movement Science Media. <https://doi.org/10.2519/jospt.2019.0702>
- 9) Kim Martineau. (2023, April 20). What is generative AI? IBM.
- 10) Malec, M. (2024, February 22). Leveraging Generative AI in the Workplace: A Guide for CXOs. Retrieved from HatchWorks AI: <https://Hatchworks.Com/Blog/Gen-Ai/Generative-Ai-Workplace-Transformation/>.

- 11) McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review*, 15(1), 28–50. <https://doi.org/10.1177/1088868310366253>
- 12) Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533–544. <https://doi.org/10.1007/s10488-013-0528-y>
- 13) Pettersson, J., Hult, E., Eriksson, T., & Adewumi, T. (2024). Generative AI and Teachers -- For Us or Against Us? A Case Study. <http://arxiv.org/abs/2404.03486>
- 14) PHREB Annual Report 2021. (2021).
- 15) Reddy, K. G., & Khan, M. G. M. (2023). Constructing efficient strata boundaries in stratified sampling using survey cost. *Heliyon*, 9(11). <https://doi.org/10.1016/j.heliyon.2023.e21407>
- 16) Su, Y., & Li, M. (2021). Applying Technology Acceptance Model in Online Entrepreneurship Education for New Entrepreneurs. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.713239>
- 17) Toyon, M. A. S. (2021). Explanatory sequential design of mixed methods research: Phases and challenges. *International Journal of Research in Business and Social Science* (2147- 4478), 10(5), 253–260. <https://doi.org/10.20525/ijrbs.v10i5.1262>
- 18) Trabelsi, M. A. (2024). The impact of artificial intelligence on economic development. *Journal of Electronic Business & Digital Economics*, 3(2), 142–155. <https://doi.org/10.1108/jebde-10-2023-0022>
- 19) Uygun, D. (2024). Teachers' perspectives on artificial intelligence in education. *Advances in Mobile Learning Educational Research*, 4(1), 931–939. <https://doi.org/10.25082/amler.2024.01.005>