Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

THE EFFECTIVENESS OF A GENERATIVE ARTIFICIAL INTELLIGENCE-BASED TRAINING PROGRAM IN DEVELOPING TEACHERS' INSTRUCTIONAL PRACTICES

AHMAD AWAD TASHTOUSH*
PhD, Philosophy of Curriculum and Teaching Methods, Faculty of Educational Sciences, Al-Hussein Bin Talal University, Ma'an, Jordan. * Corresponding Author's Email: Ahmadtashtosh1984@gmail.com

MOATH AHMAD YOUSEF
PhD, Philosophy of Curriculum and Teaching Methods, Faculty of Educational Sciences, Al-Hussein Bin Talal University, Ma'an, Jordan.

HALA SALEM HAMAD AL HJOOJ
PhD, Philosophy of Curriculum and Teaching Methods, Faculty of Educational Sciences, Al-Hussein Bin Talal University, Ma'an, Jordan.

ENSAF MUSTAFA JAMIL AL HAWAMLEH
PhD, Philosophy of Curriculum and Teaching Methods, Faculty of Educational Sciences, Al-Hussein Bin Talal University, Ma'an, Jordan.

BESSAN AHMAD KHALEEL SALEM
PhD, Philosophy of Curriculum and Teaching Methods, Faculty of Educational Sciences, Al-Hussein Bin Talal University, Ma'an, Jordan.

Abstract

This study investigated how a structured generative artificial intelligence (GenAI)-based professional development program, which builds instructional practices of teachers, is effective. The research used a quasi-experimental pretest-posttest control-group design and 80 teachers of public schools in Ma'an Governorate, Jordan were used. There were equal numbers of participants in both experimental (n = 40) and control (n = 40) groups. The intervention was eight training sessions conducted in four weeks and was based on prompt engineering, lesson planning, instructional design, differentiated learning, assessment construction, feedback generation, verification of AI outputs, and ethical and privacy considerations. A 24-item Generative AI-Enhanced Instructional Practices Scale was used to collect data, which included three dimensions: lesson planning, instructional design and teaching practices, and assessment and feedback practices. The analysis of covariance (ANCOVA) and Pearson correlation coefficients, as well as descriptive statistics, Cronbach alpha, independent-samples t tests, and analysis of covariance (ANCOVA) were performed using IBM SPSS. The findings showed acceptable-high internal consistency, baseline group equivalence, and statistically significant adjusted posttest differences in favor of the experimental group in all dimensions. Big effect sizes were realized. The results suggest that an instructional GenAI training program based on pedagogy can enhance lesson planning, instructional design, assessment, and feedback practices of teachers. The research adds a systematic intervention framework and empirical data of the public schools in Ma'an Governorate.

Keywords: Generative Artificial Intelligence; Teacher Professional Development; Instructional Practices; Lesson Planning; Assessment and Feedback; TPACK; Quasi-Experimental Research; SPSS.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

1. INTRODUCTION

Artificial intelligence has moved rapidly from a specialist technical field into mainstream educational practice, with schools increasingly using data-driven and automated tools to support teaching and teach (Chiu et al., 2023). The advent of generative artificial intelligence (GenAI), especially large language models, has expedited this shift since educators can engage with sophisticated systems by using natural-language prompts (Kasneci et al., 2023). These systems are able to write lesson plans, create learning objectives, create examples, modify explanations, create assessment items, and create feedback faster than traditional methods of preparing lessons (Lo, 2023). They have educational potential such as personalization, developing resources, formative assessment, and supporting teacher productivity (Tlili et al., 2023). Simultaneously, their application brings about issues of factual accuracy, bias, privacy, transparency, and overreliance on automated output (Farrokhnia et al., 2024).

The main educational problem is not that GenAI can produce fluent content, but that teachers can assess and use that content in a manner that enhances teaching (Miao and Holmes, 2023). The linguistic fluency might give an illusion of factual correctness and pedagogical quality since a well-polished response can still have poor arguments or unsubstantiated statements (Dwivedi et al., 2023). An AI-generated lesson can contain imprecise goals, poor sequencing, insufficient cognitive engagement, inappropriate culturally insensitive examples, or assessments that are not in line with the desired results (Kasneci et al., 2023). Big language models can also generate fake data and non-existent sources with a high level of certainty, and systematic human verification is essential (Farrokhnia et al., 2024).

The main decision makers in AI-based classrooms, however, will be teachers since they decode the standards of the curriculum, determine the needs of the learners, choose the instructional methods, and design the assessment (Celik, 2023). They also need to assess the quality of the generated explanations and decide whether a resource can be used by a particular group of learners and in a particular instructional context (Moorhouse et al., 2024). GenAI does not eliminate professional responsibility, rather, it raises the necessity to apply informed judgment regarding the situations when AI assistance is needed, how an output is to be revised, and when it is to be rejected (UNESCO, 2023). Human agency, accountability, data protection, inclusion, and critical oversight are thus cited as human-centered guidance requirements to enable responsible educational adoption (UNESCO, 2021).

The increasing access to GenAI has posed an immediate professional-development problem since the access to a tool does not guarantee the possibility of using it pedagogically (Brandão, 2024). Most educators are exposed to such systems prior to systematic training in timely building, factual validation, prejudice identification, privacy, or evaluation validity (UNESCO, 2024). The studies on the willingness of teachers show that the willingness to utilize AI and perceived usefulness do not always reflect the professional knowledge required to successfully integrate AI (Ayanwale et al., 2022). It is also necessary to have ethical and pedagogical integration, which involves knowledge beyond the general technology acceptance or confidence (Celik, 2023). As a result, the fact that teachers plan to employ GenAI cannot alone prove that they are able to construct more effective lessons or enhance the assessment practices.

Previous studies have often focused on attitudes, intentions, ethical issues, academic integrity, and potential uses of generative AI (Cotton et al., 2024). Key opportunities and threats have been identified in this literature, although much of the evidence is conceptual, descriptive, or cross-sectional (Crompton and Burke, 2023). Education ChatGPT reviews have thus requested increased intervention research that quantifies professional practice change as opposed to perceptions or intentions (Lo, 2023). This necessity is particularly evident in the research that focuses on teachers, which is not as advanced as student-centered and higher education research (Zawacki-Richter et al., 2019). Sustained, content-based, coherent, collaborative, and active learning-based professional development is more likely to impact classroom practice (Darling-Hammond et al., 2017). Short demonstrations can present a new tool, yet they offer few possibilities to teachers to test outputs, find errors, update materials, and relate technological functions to the curriculum and pedagogy (Brandão, 2024). Good programs must also provide a chance to practice new knowledge to real world work and get feedback on performance (Desimone, 2009). These features are especially significant to GenAI since professional competence is built up through recurrent assessment and correction instead of being exposed to the features of the tools in isolation (Al-Ali and Miles, 2025).

The current research was based on four mutually supporting views. First, GenAI was considered an educational affordance that is valued based on the educational intent, task requirements, and the quality of human supervision (Chiu et al., 2023). Second, AI literacy was perceived as the ability to perceive, apply, assess, and interact with AI systems in an ethical manner (Ng et al., 2021). Third, Technological Pedagogical Content Knowledge framework was adopted to clarify the need to integrate technological knowledge with pedagogy and subject matter as opposed to developing it in isolation (Mishra and Koehler, 2006). Fourth, the explanations of how guided practice, reflection, feedback and mastery experiences could help in supporting professional change were explained in terms of experiential learning and teacher self-efficacy (Bandura, 1997; Kolb, 1984).

In this context, the researchers tested a four-week GenAI-based training program on lesson planning, instructional design and teaching practices, and assessment and feedback practices. These domains represent key domains where teachers can implement GenAI and still have professional control of curricular alignment and the quality of instruction (Moorhouse et al., 2024). Pretest-posttest control-group design allowed the study to compare the groups and consider the initial differences by analyzing the covariance. This method is a reaction to recent demands to provide more evidence regarding teacher-centered GenAI professional development (Tan et al., 2025). It also goes beyond conclusions that are based on attitudes only and looks into quantifiable differences after a systematic intervention (Al-Ali and Miles, 2025).

The Jordanian setting provides the study with further relevance since educational systems are seeking to transform digitally, whereas access, infrastructure, previous

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

experience, and professional assistance is not evenly distributed across the locations (OECD, 2023). Structured professional development may be used in situations where there is an uneven support to differentiate between superficial adoption and pedagogically responsible integration (UNESCO, 2024). The emphasis on teachers in Ma’an Governorate adds to the body of evidence in a school environment that is not technologically advanced and higher education-oriented, which is prevalent in the majority of the global AI literature (Crompton and Burke, 2023). This local attention is also concerned with the underrepresentation of school teachers in previous studies of artificial intelligence in education (Zawacki-Richter et al., 2019).

1.1 Statement of the Problem

Despite rapid access to generative AI, many teachers have not received systematic preparation for its pedagogical use (UNESCO, 2024). Informal experimentation may privilege speed and novelty over curricular alignment, cognitive demand, factual verification, inclusion, privacy, and assessment validity (Miao & Holmes, 2023). Without professional preparation, GenAI may automate weak instructional practices instead of improving them (Celik, 2023). The research problem is therefore the limited intervention evidence showing whether sustained, pedagogically grounded training can produce measurable improvements in teachers’ instructional practices (Tan et al., 2025). This evidence gap is particularly relevant in underrepresented school contexts such as Ma’an Governorate, where local empirical research can complement findings from higher education and technology-rich settings (Crompton & Burke, 2023).

1.2 Purpose of the Study

This study was aimed at identifying how well a generative artificial intelligence-based training program can be used to build the instructional practices of teachers. In particular, the researchers tested the hypothesis that teachers who participated in the program had better posttest scores compared to teachers in a control group when pretest scores were statistically adjusted.

1.3 Objectives

1) To design a structured professional development program that integrates generative AI with pedagogical planning, instruction, assessment, verification, and ethical practice.
2) To measure teachers’ GenAI-enhanced lesson-planning practices before and after the intervention.
3) To measure teachers’ GenAI-enhanced instructional-design and teaching practices before and after the intervention.
4) To measure teachers’ GenAI-enhanced assessment and feedback practices before and after the intervention.
5) To estimate the effect of the training program using ANCOVA and report practical effect sizes.

ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

1.4 Research Hypotheses

1) H01: There is no statistically significant difference at α = .05 between the adjusted posttest mean scores of the experimental and control groups in lesson planning.
2) H02: There is no statistically significant difference at α = .05 between the adjusted posttest mean scores of the experimental and control groups in instructional design and teaching practices.
3) H03: There is no statistically significant difference at α = .05 between the adjusted posttest mean scores of the experimental and control groups in assessment and feedback practices.
4) H04: There is no statistically significant difference at α = .05 between the adjusted posttest mean scores of the experimental and control groups in overall GenAI-enhanced instructional practices.

1.5 Significance of the Study

It has theoretical, methodological and practical value in the study. In theory, it relates the integration of generative AI with TPACK, AI literacy, experiential learning and teacher self-efficacy. The methodologically, it provides a manageable intervention design, which can be analyzed with SPSS without subjecting the research question to a wrong structural model. In practice, it offers a replicable professional-development sequence, a measurement instrument, and an analysis plan which can be adapted by schools and researchers.

1.6 Conceptual and Operational Definitions

Generative Artificial Intelligence

Conceptual Definition: Is artificial intelligence systems that produce new content, like text, images, audio, or code, in response to user prompts using large language models and other machine-learning methods (Kasneci et al., 2023; UNESCO, 2023).

Operational Definition: In this study, generative artificial intelligence is the language-model-based tool that teachers will use in the training program to aid lesson planning, design of instructional activities, construction of assessment, preparation of feedback, and evaluation and revision of generated educational materials.

Genai-Based Training Program

Conceptual Definition: Is a professional-development intervention that is structured to build knowledge and skills of teachers in pedagogical, critical, and responsible use of generative artificial intelligence in education (Holmes et al., 2019; UNESCO, 2023).

Operational Definition: Is an independent variable in this study. It was a series of eight training sessions in four weeks to the teachers in the experimental group. The sessions were based on theoretical explanation, modeling by the trainer, guided practice, individual practice, peer review, verification of the output and reflective activities.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

Teachers' Instructional Practices

Conceptual Definition: The instructional practices of teachers are defined as the professional actions, strategies and decisions that teachers use in planning, implementing, assessing and evaluating instruction in order to facilitate effective student learning (Mishra and Koehler, 2006; Shulman, 1986).

Operational Definition: Teachers instructional practices in this study are the dependent variable that will be measured using the Generative AI-Enhanced Instructional Practices Scale. The scale was made up of 24 items that were spread out in three dimensions; lesson-planning practices, instructional-design and teaching practices, and assessment and feedback practices. The total score was obtained as the average of all items, where the higher the score, the more the generative AI is integrated in the instructional practice of teachers.

2. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

2.1 Generative AI as A Pedagogical Affordance

Generative artificial intelligence differs from conventional educational technologies because it can produce new text, examples, explanations, questions, scenarios, and feedback in response to natural-language prompts (Kasneci et al., 2023). This generative ability enables educators to have quick access to alternative teaching materials and representations (Baidoo-Anu & Owusu Ansah, 2023). Nonetheless, the speed of generation does not imply the factual accuracy, alignment with the curriculum, and quality of pedagogy (Farrokhnia et al., 2024). Big language models generate statistically plausible answers instead of education that is independently validated (Dwivedi et al., 2023). This means that teachers have to assess, modify, and even discard the outputs prior to their use in the classroom (UNESCO, 2023).

Automatic content production is not the most powerful educational application of GenAI, but assisted pedagogical design (Miao and Holmes, 2023). It can help teachers brainstorm quantifiable goals, detect misconceptions, and develop various explanations of challenging concepts (Lo, 2023). They also have the ability to produce differentiated versions of the tasks, initial assessment items, and draft feedback that can be reviewed by a professional later (Tlili et al., 2023). Such functions can decrease the amount of routine work and increase the scope of the available resources (Grassini, 2023). Their worth, however, is conditional on the explicit contextual limitations and systematic comparison to the curriculum outcomes and the needs of learners (Moorhouse et al., 2024).

There are also risks made by GenAI that have a direct impact on the quality of instruction. When the outputs are not verified, hallucinated facts and fabricated references may find their way into learning materials (Kasneci et al., 2023). Cultural and linguistic bias can result in the creation of unsuitable examples or support unequal representation (UNESCO, 2021). Weak distractors, ambiguous wording, or false keys can be generated assessment items (Farrokhnia et al., 2024). Automated feedback can be fluent and

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

generic or not sensitive to the real difficulty of the learner (Lim et al., 2023). Privacy breaches may happen when the identifiable learner data is typed into the tools without the agreed protection (UNESCO, 2023).

This paper therefore theorizes GenAI as cognitive and design support and not an independent teacher. Its pedagogical usefulness is likely to manifest itself in a human-in-the-loop process of identifying the instructional problem, creating an output, reviewing evidence, updating the material, and assessing its appropriateness (OECD, 2023). This cycle maintains teacher agency since the ultimate decision on instruction lies with the teacher (Celik, 2023). It also promotes accountability as it demands that teachers explain how created content was verified and modified (UNESCO, 2024). This process still requires transparency, inclusion, and data protection instead of considering them as optionalities once the content is generated (UNESCO, 2021).

2.2 Teacher AI Literacy and Professional Competence

AI literacy is the knowledge, skills, attitudes, and ethical awareness needed to comprehend, utilize, assess, and critically interact with artificial intelligence (Ng et al., 2021). It involves being aware of what AI systems are capable of and cannot do and that their results are likely to be probabilistic, incomplete, or biased (Long and Magerko, 2020). It also involves making informed decisions regarding the tasks that AI support is suitable and those that need to be performed by humans directly (UNESCO, 2024). In the case of teachers, AI literacy goes as far as curricular alignment, pedagogical appropriateness, assessment validity, learner protection, and professional responsibility (Celik, 2023).

Teacher AI competence goes beyond confidence, positive attitudes, and frequency of use (Ayanwale et al., 2022). GenAI can be utilized regularly by a teacher who still cannot detect misinformation and safeguard student information (UNESCO, 2024). Repeated use is also not a guarantee that an activity generated by AI is beneficial to meaningful learning or that it has the right cognitive demand (Kasneci et al., 2023). The competence should thus be reflected in the quality of professional judgments and teaching artifacts (Moorhouse et al., 2024). The informed trust that should be developed as a result of professional learning should not be blind trust or complete distrust towards the output of AI (Nazaretsky et al., 2022).

The training program builds up four applied competences. Timely construction involves the teacher specifying context, audience, constraints, examples, and the format of the output that is needed (Al-Ali and Miles, 2025). Pedagogical adaptation involves matching the response generated with the learning outcomes, the characteristics of the learner, and the disciplinary content (Celik, 2023). Factual accuracy, bias, cognitive demand, and quality of assessment are the aspects of critical verification that should be examined (Miao and Holmes, 2023). Privacy protection, transparency, and maintenance of human responsibility towards educational consequences is essential in ethical decision-making (UNESCO, 2024).

Calibrated trust is also needed in professional competence. The distrust in all the AI outputs can make teachers overlook the useful types of assistance, and the overtrusting

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

of the output can make them accept the inaccurate or pedagogically inadequate material (Nazaretsky et al., 2022). The basis of conditional trust must be the task, the evidence available, institutional policy and the cost of error (UNESCO, 2023). Tasks with high stakes like grading or decisions involving individual learners need a higher level of verification compared to low-stakes brainstorming (Miao and Holmes, 2023). Professional learning that is structured can assist teachers in becoming more critical of AI recommendations and can be more responsible in their use (Nazaretsky et al., 2022).

2.3 TPACK and Pedagogical Integration of Genai

According to Technological Pedagogical Content Knowledge framework, the key to successful technology integration lies in the interaction between the technological knowledge, pedagogical knowledge and content knowledge (Mishra and Koehler, 2006). Technological knowledge deals with the abilities, restrictions and functionality of the tool (Koehler and Mishra, 2009). Pedagogical knowledge helps teachers to assess the learning processes, differentiation, classroom interaction, and the design of assessment (Celik, 2023). Content knowledge helps them to detect disciplinary errors and maintain conceptual integrity of resources generated (Shulman, 1986). The integration of GenAI is thus undermined when one of these areas of knowledge is considered separately.

The interaction of the knowledge domains of TPACK is particularly apparent in GenAI. Even a technically competent teacher can produce weak instruction by writing advanced prompts and failing to meet the objectives that are not clear or with misaligned activities (Celik, 2023). A teacher who has not had a pedagogical experience and has limited knowledge of AI might not be able to convey enough constraints to the system or assess its technical constraints (UNESCO, 2024). To be used effectively, the teacher should bridge the gap between system affordances and subject-specific goals and prior knowledge of learners (Moorhouse et al., 2024). It also involves coordination of teaching strategies, learning activities and demonstration of success (Mishra and Koehler, 2006).

The training program realizes the TPACK through the structuring of each AI task in terms of a real-life instructional problem. Participants are trained to have subject-specific goals instead of training on isolated chatbot commands (Al-Ali and Miles, 2025). They create and criticize differentiated activities concerning the needs of learners (Moorhouse et al., 2024). They build assessment blueprints and check answer keys that are generated prior to accepting items (Farrokhnia et al., 2024). They also update feedback to maintain accuracy, clarity and teacher accountability (Miao and Holmes, 2023). Technology in this manner is incorporated in pedagogical rationale and content validation as opposed to being considered as a separate innovation.

This theoretical stance describes the choice of the dependent variables. Lesson planning is the organization of the content, goals, sequence and teaching materials (Mishra and Koehler, 2006). Instructional design can be considered the conversion of pedagogical and content knowledge into the activities that facilitate engagement and meaningful learning (Celik, 2023). Assessment gives a testament of learning and must be in tandem with the desired results (Black and Wiliam, 1998). Timely, specific and linked to clear

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

performance goals are the characteristics of feedback that promote improvement (Hattie and Timperley, 2007). Combined, these dimensions offer a more justifiable gauge of GenAI integration as compared to general intention to use the technology (Tan et al., 2025).

2.4 Professional Development, Experiential Learning, and Teacher Self-Efficacy

Teacher professional development is effective, content-oriented, active, collaborative, and related to classroom practice (Darling-Hammond et al., 2017). Single demonstrations can raise awareness but seldom give enough time to experiment and refine professional work (Brandão, 2024). Active learning enables teachers to implement a new method to real tasks rather than passively receive information (Desimone, 2009). Coherence bridges the gap between the program and the curriculum, school expectations and existing knowledge of teachers (Garet et al., 2001). Professional learning is more likely to impact practice when it is followed up and provided with feedback opportunities (Guskey, 2002).

The form of the intervention is well justified with the help of experiential learning. The first step is that teachers are presented with a tangible instructional problem and apply GenAI to generate a preliminary response (Kolb, 1984). They then examine the strengths and weaknesses of such output and consider why it was so good. General principles are based on the experience and implemented on a revised prompt or instructional product. New GenAI professional-development interventions also focus on applied tasks, iterative experimentation, and reflection over passive tool demonstrations (Al-Ali and Miles, 2025). This cycle is supposed to build practical professional knowledge since those involved in it will repeatedly relate conceptual guidance and real instructional products.

The concept of teacher self-efficacy is applicable since teachers will initiate and continue with new practices when they are convinced that they can do it well (Bandura, 1997). Modeling will demonstrate to participants how an expert user can assess and re-edit AI output (Al-Ali and Miles, 2025). Structured mastery experiences are offered by guided practice and feedback, which minimise uncertainty in initial use (Nazaretsky et al., 2022). Peer comparison has the potential to expose the teachers to alternative strategies and normalization of revision of weak outputs. The goal, though, is calibrated self-efficacy: educators must be confident in their ability to use the technology and have enough critical awareness to understand when the output of the technology is not satisfactory (UNESCO, 2024).

The professional-development model also presupposes that the change in beliefs becomes more stable in case the teachers see improvement in their own professional performance (Guskey, 2002). This is why participants produce physical products, such as lesson plans, differentiated activities, assessment packages, and examples of feedback. Every product is evaluated in terms of clear quality standards and not in terms of newness or rapidity. Visible artifacts also enable participants to contrast the initial and revised work, enhancing reflection and professional responsibility (Al-Ali and Miles, 2025).

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

The four theoretical axes combined contribute to the conceptual model of the study. GenAI offers quick access to other content and teaching models, yet its educational worth relies on the critical human supervision (Kasneci et al., 2023). AI literacy builds the ability to comprehend, assess, and utilize these systems in a responsible manner (Ng et al., 2021). TPACK describes the way that technological affordances should be aligned with pedagogy and subject knowledge (Mishra and Koehler, 2006). Repeated application, reflection, and revision are offered through experiential professional learning (Al-Ali and Miles, 2025). Such mechanisms will reinforce the lesson planning, instructional design and assessment and feedback practices once the baseline performance is statistically controlled.

3. METHODOLOGY

3.1 Research Approach and Design

A quantitative quasi-experimental pretest- posttest control-group design was used to conduct the study. The experimental group was assigned to the GenAI-based training program, and the control group to the traditional activities of professional-development. Both groups took the same measure both prior and after the intervention. ANCOVA was chosen as the most important analysis since it compares post test scores and statistically controls the performance in pretests.

3.2 Participants

The sample of the study comprised of 80 teachers in public basic and secondary schools in Maan Governorate, Jordan. The number of teachers in the experimental group was 40 and the number of teachers in the control group was 40. The available schools that were willing to participate were recruited through an easy sampling method. The existing teacher groups were maintained to ensure that the school schedules were not disrupted; therefore, the study was categorized as quasi-experimental and not totally random. The sample sample consisted of the teachers of various academic backgrounds, teaching experience, level of school, and the degree of prior exposure to generative AI. Table 1 summarizes the demographic and professional features of the teachers who participated.

Table 1: Characteristics of the Study Participants

Variable	Category	n
Group	Control / Experimental	40 / 40
Gender	Male / Female	34 / 46
Qualification	Bachelor / Master / Doctorate	56 / 19 / 5
School level	Primary / Secondary	45 / 35
Prior AI experience	No / Yes	46 / 34

As indicated by Table 1, the sample was balanced in terms of the experiment and control groups and the participants were represented by both genders, various academical qualifications, basic and secondary school level, and varying levels of prior exposure to generative AI.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

3.3 Variables

The study variables are as shown in Table 2 along with their roles in the research design and the operationalization of each variable is presented.

Table 2: Study variables and operational definitions

Role	Variable	Operationalization
Independent variable	Participation in the GenAI-based training program	0 = Control; 1 = Experimental
Dependent variable 1	Lesson-planning practices	Mean of Items 1-8
Dependent variable 2	Instructional-design and teaching practices	Mean of Items 9-16
Dependent variable 3	Assessment and feedback practices	Mean of Items 17-24
Overall dependent variable	Overall GenAI-enhanced instructional practices	Mean of Items 1-24
Covariate	Corresponding pretest score	Continuous score from 1 to 5

As shown in Table 2, the independent variable was participation in the training program based on GenAI, the three dimensions of instruction/practice and the overall score were the dependent variables, and the pretest score was used as a covariate in each ANCOVA model.

3.4 The Training Program

The procedure was carried out during four weeks in a row and comprised eight two-hour sessions. The program was modeled on the basis of the real teacher activities and the cycles of demonstration, guided practice, and review by peers, independent revision, and reflection. The structured evaluation checklist used enabled the participants to analyze accuracy, alignment, cognitive level, inclusion, bias, and privacy and teacher responsibility. Table 3 outlines the timing, contents and anticipated output of the eight training sessions.

Table 3: Content of the Genai-Based Training Program

Session	Theme	Main content	Required product
1	GenAI foundations	Capabilities, limitations, hallucination, bias, privacy, and the teacher's role	Risk-analysis worksheet
2	Prompt design	Context, role, audience, constraints, examples, output format, and iterative prompting	Prompt improvement log
3	Learning objectives	Measurable objectives, curriculum alignment, and prerequisite knowledge	Objective-alignment task
4	Lesson planning	Sequencing, timing, resources, active learning, and verification	Complete lesson-plan draft
5	Instructional design	Scaffolding, questioning, multiple explanations, and cognitive demand	Redesigned activity
6	Differentiation and inclusion	Readiness, accessibility, language support, and stereotype review	Differentiated task set
7	Assessment and feedback	Blueprints, item quality, rubrics, answer verification, and feedback	Assessment package
8	Integration and reflection	Ethical disclosure, implementation plan, peer critique, and final revision	Final portfolio

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

As demonstrated in Table 3, the program advanced to the initial knowledge and immediate design to lesson planning, instructional design, differentiation, assessment, feedback, and overall integrated ethical use. The sessions needed a tangible professional product so that training was not based on theoretical presentation, but practice.

3.5 Training Fidelity

The fidelity to implementation was observed by attendance, completion of necessary tasks, trainer checklists, participant prompt logs and reviewing session artifacts. The minimum exposure criterion was to attend at least 80% of the sessions. The trainer recorded any variations of the scheduled order or time of the program.

3.6 Instrumentation

The Generative AI-Enhanced Instructional Practices Scale was 24 items with 1 (strongly disagree) to 5 (strongly agree) rating scales. Lesson planning was measured with eight items, instructional design and teaching with eight, and assessment and feedback with eight. The mean of the relevant items was used to obtain dimension scores. The total score was the average of 24 items.

3.7 Content Validity

The scale and training materials were reviewed by a panel of specialists in educational technology, curriculum and instruction, measurement and evaluation, teacher professional development, and AI ethics. The reviewers assessed the applicability, understandability, thoroughness, and correspondence of the items to the operational definitions. Based on their comments, wording was revised, ambiguity eliminated and the fit between the instrument dimensions and the study variables was further enhanced.

3.8 Pilot Testing and Reliability

A pilot study was carried out among a sample of the teachers who did not form the final study sample. The pilot test involved the evaluation of the clarity of items, time taken to respond, variation in responses, and internal consistency. Some slight linguistic changes were also done according to the feedback of the participants and suggestions of the specialists to enhance the clarity of some items and then the final version of the instrument was given. The alpha coefficients of Cronbach were then obtained in each dimension and the overall scale at pretest and posttest and the obtained Cronbach alpha values justified the reliability of the instrument.

3.9 Data Collection Procedures

1) Prior to data collection, institutional ethical approval and permission was obtained by the participating schools.
2) The teachers were sampled among attending schools in the Ma’an Governorate and all the teachers were informed and gave their consent.
3) Both groups were given the pretest questionnaire and demographic data was gathered prior to the intervention.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

4) The experimental group had four-week program of GenAI-based training, and the control group had regular professional-developmental programs.
5) Attendance, implementation fidelity, and participant artifact records were maintained throughout the intervention.
6) The posttest questionnaire was administered to both groups under comparable conditions immediately after the training program was completed.
7) The data collected were coded, entered, screened, cleaned and analyzed with the use of IBM SPSS.

3.10 Ethical Considerations

Data collection was preceded by the acquisition of ethical approval and administrative permission. All participants were allowed to participate voluntarily, and informed consent was obtained. Anonymous coding and the secure data storage ensured secrecy and the participants had the right to withdraw without any punishment.

Employment evaluation was not related to participation. Educators were advised to avoid inputting recognizable information on students in publicly available GenAI tools and the training focused on adherence to institutional privacy and responsible-use policies.

3.11 Statistical Analysis

IBM SPSS was used to analyze the data. Frequencies and descriptive statistics were used for data screening. The internal consistency was evaluated using Cronbachs alpha and independent-samples t tests were used to test the equivalence at baseline. Pearson correlation coefficients were determined to test the relationships that exist between the three instructional-practice dimensions.

Each hypothesis was run individually with analysis of covariance (ANCOVA) as the dependent variable was the corresponding posttest score, the fixed factor was group, and the covariate was the corresponding pretest score. Independence, linearity, homogeneity of regression slopes, normality of residue and homogeneity of variance were all assumed. The measure of effect-size was reported to be partial eta squared.

4. RESULTS

4.1 Preliminary Analyses

The hypothesis testing was preceded by the screening of the dataset. All the answers were full and within the allowable range of the five-point scale. Before interpreting the effects of the interventions, internal consistency, baseline equivalence, intercorrelations between the study dimensions, and assumptions needed to run ANCOVA were analyzed.

The internal-consistency coefficients for the scale dimensions and total score at pretest and posttest are reported in Table 4.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

Table 4: Internal-Consistency Coefficients for the Study Scale

Wave	Scale	Items	Cronbach alpha
Pre	Lesson planning	8	0.848
Pre	Instructional design	8	0.888
Pre	Assessment and feedback	8	0.825
Pre	Overall scale	24	0.741
Post	Lesson planning	8	0.869
Post	Instructional design	8	0.903
Post	Assessment and feedback	8	0.856
Post	Overall scale	24	0.887

The coefficients of Cronbach alpha as indicated in Table 4 varied between.741 and.903. These values demonstrate that the three dimensions and the overall scale have an acceptable to high internal consistency, which can justify the usage of the obtained scores in the further hypothesis testing. Independent-samples t tests were used to test the hypothesis of the similarity of the two groups prior to the intervention, as shown in Table 5.

Table 5: Baseline Equivalence of the Experimental and Control Groups

Measure	Experimental Group		Central Group		t	p
	M	SD	M	SD		
Lesson planning	2.99	0.60	3.12	0.81	-0.81	0.423
Instructional design	3.07	0.80	2.85	0.73	1.29	0.200
Assessment and feedback	3.12	0.59	3.02	0.72	0.72	0.474
Overall	3.06	0.37	2.99	0.40	0.78	0.440

The analysis in Table 5 shows that all the differences in pretest lesson planning, instructional design, assessment and feedback, and the overall score were non-significant (all $p > .05$).

The two groups thus started the intervention with similar measured levels although ANCOVA was still used as a means of counteracting the residual variation at the baseline. Table 6 shows the relationship between the posttest instructional-practice dimensions.

Table 6: Pearson Correlations among Posttest Instructional-Practice Dimensions

Variable	Lesson planning	Instructional design	Assessment and feedback	Overall practices
Lesson planning	1.000	0.209	0.151	0.623
Instructional design	0.209	1.000	0.439	0.792
Assessment and feedback	0.151	0.439	1.000	0.728
Overall practices	0.623	0.792	0.728	1.000

The three dimensions had a positive correlation as indicated in Table 6. The most significant interdimensional correlation was between instructional design and assessment and feedback ($r = .439$), whereas the correlation with the overall score was larger since the total score was calculated using the same component dimensions. The findings of diagnostic tests to assess the key ANCOVA assumptions are given in Table 7.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

Table 7: Diagnostic Tests for the ANCOVA Assumptions

Outcome	Slope interaction p	Shapiro p	Levene p
Lesson planning	0.159	0.887	0.555
Instructional design	0.098	0.405	0.315
Assessment and feedback	0.639	0.034	0.267
Overall instructional practices	0.549	0.112	0.062

The group-by-pretest interaction was nonsignificant in all outcomes, which indicates that the regression slopes are homogeneous, and all Levene tests are nonsignificant as shown in Table 7. Despite the significance of the Shapiro-Wilko test to assess and provide feedback, the analysis of the residual distribution and equal sizes of the groups did not suggest a serious infraction; thus, ANCOVA was not rejected.

4.2 Testing Hypothesis H01: Lesson Planning

H01: There is no statistically significant difference at α = .05 between the adjusted posttest mean scores of the experimental and control groups in lesson planning.

Table 8 shows the observed pretest and posttest means and the covariate-adjusted posttest means of lesson planning.

Table 8: Descriptive and adjusted means for lesson planning

Group	Pretest M	Posttest M	Adjusted posttest M
Control	3.12	3.11	3.07
Experimental	2.99	3.86	3.89

As Table 8 indicates, the mean in the experimental group rose to 3.86 and 2.99 in pretest and posttest respectively, but in the control group, there was no significant change. The experimental-group mean (adjusted by the baseline scores) was 3.89 versus 3.07 in the control group. Table 9 presents the ANCOVA results that are used to test H01.

Table 9: ANCOVA results for lesson planning

Source	F	df1	df2	p	Partial eta squared
Group	45.69	1	77	< .001	0.372

As The adjusted group effect on lesson planning as shown in Table 9 was significant, $F(1, 77) = 45.69$, $p < .001$, partial $\eta^2 = .372$. As a result, H01 was disproved, and the outcome was in favor of teachers who underwent the training program based on GenAI.

4.3 Testing Hypothesis H02: Instructional Design and Teaching Practices

H02: There is no statistically significant difference at α = .05 between the adjusted posttest mean scores of the experimental and control groups in instructional design and teaching practices.

The observed and adjusted means for instructional design and teaching practices are presented in Table 10.

ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

Table 10: Descriptive and Adjusted Means for Instructional Design and Teaching Practices

Group	Pretest M	Posttest M	Adjusted posttest M
Control	2.85	3.01	3.07
Experimental	3.07	4.05	3.99

Table 10 shows that the experimental group that started with a mean of 3.07 in the pretest actually improved to 4.05 in the posttest as compared to the control group which only improved to 3.01. The adjusted posttest means of the experimental group were 3.99 and that of the control group was 3.07. Table 11 presents the ANCOVA results that were used to test H02.

Table 11: ANCOVA Results for Instructional Design and Teaching Practices

Source	F	df1	df2	p	Partial eta squared
Group	68.00	1	77	< .001	0.469

As presented in Table 11, the adjusted group effect was statistically significant, $F(1, 77) = 68.00$, $p < .001$, partial $\eta^2 = .469$. H02 was thus dismissed which means that there was a significant intervention effect on the instructional design and teaching practices.

4.4 Testing Hypothesis H03: Assessment and Feedback

H03: There is no statistically significant difference at $\alpha = .05$ between the adjusted posttest mean scores of the experimental and control groups in assessment and feedback practices.

The observed and adjusted means for assessment and feedback practices are presented in Table 12.

Table 12: Descriptive and adjusted means for assessment and feedback

Group	Pretest M	Posttest M	Adjusted posttest M
Control	3.02	3.12	3.14
Experimental	3.12	4.05	4.03

Table 12 indicates that the mean of the experimental group rose to 4.05 at the end of the experiment (posttest) and 3.12 at the beginning of the experiment (pretest), whereas the mean of the control group only rose to 3.12 at the end of the experiment (posttest) and 3.02 at the beginning of the experiment (pretest). The mean of the experimental group and that of the control group after adjustment was 4.03 and 3.14 respectively. Table 13 shows the ANCOVA results that are used to test H03.

Table 13: ANCOVA Results for Assessment and Feedback

Source	F	df1	df2	p	Partial eta squared
Group	65.27	1	77	< .001	0.459

The adjusted difference between the groups was statistically significant as shown in Table 13, $F(1, 77) = 65.27$, $p < .001$, partial $\eta^2 = .459$. Based on this, H03 was rejected with the difference going in the favor of the experimental group.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

4.5 Testing Hypothesis H04: Overall Instructional Practices

H04: There is no statistically significant difference at $\alpha = .05$ between the adjusted posttest mean scores of the experimental and control groups in overall GenAI-enhanced instructional practices.

The overall observed and adjusted instructional-practice means are presented in Table 14.

Table 14: Descriptive and Adjusted Means for Overall Instructional Practices

Group	Pretest M	Posttest M	Adjusted posttest M
Control	2.99	3.08	3.09
Experimental	3.06	3.99	3.98

Table 14 indicates that the experimental group improved in pretest of 3.06 to 3.99 in the posttest as compared to a lesser improvement in the control group of 2.99 to 3.08. The posttest means adjusted were 3.98 and 3.09, respectively. The results of ANCOVA to test H04 are in Table 15.

Table 15: ANCOVA Results for Overall Instructional Practices

Source	F	df1	df2	p	Partial eta squared
Group	211.10	1	77	< .001	0.733

According to Table 15, the overall group effect adjusted was significantly different, $F(1, 77) = 211.10$, $p = .001$, partial $\eta^2 = .733$. H04 was thus discarded. This was the highest impact that was witnessed, and this suggested an improvement in the integrated instruction-practice dimensions.

5. DISCUSSION

5.1 Discussion of H01: Lesson Planning

The initial hypothesis was related to the impact of the training program based on the GenAI on the lesson-planning of teachers. The experimental group had a posttest adjusted mean of 3.89, as compared to the control group who had a posttest adjusted mean of 3.07 (Table 8 and Table 9). The statistically significant result of the ANCOVA was $F(1, 77) = 45.69$, $p < .001$, with a partial eta squared =.372. This effect size shows that group membership has explained a significant percentage of the adjusted variance in posttest lesson-planning scores when the pretest lesson-planning scores were adjusted. The numerical pattern thus indicates that the difference was not just a small increase in the experimental group, this was an advantage that was actually meaningful compared to the control condition.

One of the logical reasons why this outcome could be expected is that lesson planning is one of the spheres where the immediate and visible assistance of GenAI could be offered. A conversational system allows teachers to generate alternative objectives, examples, sequences of activities, questions and adaptations in a very short time. The initial generated response, however, was not considered to be a complete lesson by the

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

program. Students were taught to indicate the curriculum outcome, the level of the learner, the amount of time available, the instructions and limitations of the instruction, and the output format needed and to review the answer with regard to accuracy, coherency and alignment. This iterative approach probably led to better quality in the planning decisions of teachers since it involved fast generation of ideas and formal professional feedback (Kasneci et al., 2023; Miao and Holmes, 2023).

The finding also implies that the intervention increased the ability of teachers to shift towards pedagogically purposeful prompting as opposed to generic prompting. In practice, the effectiveness of GenAI in planning the lesson process is greatly dependent on the quantity and quality of background information that is offered by the teacher. Prompts that define the learning goal, students, desired level of thinking, evidence in the assessment and classroom limitations have more chances of generating workable drafts as compared to short requests on a complete lesson. Recent publications on the professional GenAI competence of teachers also highlight that to use AI effectively, teachers need to critique, modify, and defend AI-generated texts, not just to create them (Moorhouse et al., 2024; Tan et al., 2025).

This result is consistent with Al-Ali and Miles (2025), whose five-week professional-development program revealed that tailored training can assist teachers in becoming more inclined and responsible to integrate GenAI into the teaching process. It is also aligned with Li, Wang, and Bonk (2026), who discovered that the K-12 teachers were widely using ChatGPT as a source of ideas and structures of pedagogies, classroom activities, and lesson plans. The current finding builds on them by demonstrating a statistically adjusted difference between a trained group and a comparison group as opposed to basing the results merely on perceptions or descriptions of use.

The outcome must not be construed literally though. Recent articles have cautioned that GenAI is capable of producing refined yet generalized plans, inaccurate content, and poor sequencing, as well as poorly aligned activities with local curricula (Farrokhnia et al., 2024; UNESCO, 2024). As such, the improvement which has been observed is more likely to be the product of GenAI support, pedagogic standards, checking, and reworking than the technology itself. This difference is significant as it may lead to the possibility of producing more planning materials without enhancing their educational value in case of using them blindly.

5.2 Discussion of H02: Instructional Design and Teaching Practices

The second hypothesis explored the teaching practices and instructional design. As demonstrated in Tables 10 and 11, the experimental group had an adjusted posttest mean of 3.99, as compared to 3.07 in the control group. The ANCOVA result, $F(1, 77) = 68.00$, $p < .001$, partial $\eta^2 = .469$, indicates a large intervention effect. The group effect accounted for almost half of the adjusted variance of the tested model, indicating that the training was closely related to the ability of teachers to design, adapt, and evaluate AI-supported instructional activities.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

The given outcome could be attributed to the implemented design of the program. Participants trained in multiple explanation generations, active-learning exercises, cooperative exercises, differentiated materials and cognitively challenging questions. They had to also make comparisons between alternatives and alter outputs that were too generic, too simple, culturally unsuitable or not well matched to the desired learning outcome. The training was rather instructive decision-making and not content generation. This practice would be aligned with the idea of the AI competence that has been introduced recently and oriented at the critical use, pedagogical adaptation, and ethical judgment on top of the technical competence (Celik, 2023; UNESCO, 2024).

The outcome also suggests that teachers found it helpful to use GenAI as the source of alternatives but not as the alternative to instructional design. One big benefit of generative systems is that they can generate multiple representations, examples, scenarios or task versions at high speed. This may be used to support differentiation whereby teachers input information regarding readiness, level of language, learning needs and instructional objectives. However the teacher will need to decide whether the options maintain conceptual fidelity and meaningful intellectual challenge. The fact that the improvement was in the experimental group indicates that the intervention enabled participants to exercise this evaluative role in a more systematic manner (Chiu et al., 2023; Moorhouse et al., 2024).

The conclusion aligns with Moorhouse et al. (2024), whose intervention in language-teacher education yielded an increase in professional GenAI competence achieved by guided and context-specific application. It also concurs with Tan et al. (2025), whose systematic review has found that teacher professional development is at the core of bridging the difference between the ever-increasing need to integrate AI and the lack of structured preparations. The customized training and pedagogical support were also highlighted by Al-Ali and Miles (2025) as the prerequisites to an effective integration.

Simultaneously, the outcome is opposite to apprehensions that GenAI is going to streamline instructional design or push teachers to promote empty activities. These concerns hold when we have generated outputs that are not subjected to critique. Recent research has reported the concerns of teachers concerning accuracy, student dependence, academic integrity, and grade-/subject-specific guidance (Chen et al., 2025; Farrokhnia et al., 2024). The existing conclusion does not eliminate those concerns, on the contrary, it implies that it is possible to mitigate part of the risks by means of structured training and transforming evaluation, adjustment and rejection of weak outputs into explicit aspects of professional practice.

5.3 Discussion of H03: Assessment and Feedback

The third hypothesis was on assessment and feedback practices. Tables 12 and 13 indicate that the adjusted posttest mean of the experimental group and the control group was 4.03 and 3.14 respectively. The group effect was statistically significant, $F(1, 77) = 65.27$, $p < .001$, with partial $\eta^2 = .459$. This is a big effect and means that the training was related to a significant increase in the ability of teachers to produce, analyze, and revise

AI-based assessment and feedback materials once the differences in pretests are eliminated.

An especially challenging field of GenAI application is assessment since mistakes might have a direct impact on conclusions regarding student learning. The program was thus not confined to the creation of questions. The participants were collaborating on assessment blueprints, learning-outcome alignment, level of cognition, quality of distractors, check of answer keys, and criteria of rubrics, specificity of feedback, reviewing bias, and privacy. This scope can give a decent explanation of the good outcome. Teachers were not conditioned to generate more assessment content; they were conditioned to assess the validity, accuracy, fairness and usability of content that was generated.

The outcome can also be explained concerning the speed and flexibility of GenAI. A teacher may solicit multiple versions of a question, alternative feedback statements or rubric descriptors and then choose or change the best one. This can minimize the preparation routine requirements and provide additional formative assessment opportunities. Nevertheless, the generated assessment material can be ambiguous, have wrong answers, lack consistency in difficulty, or be misleading. Human evaluation is thus a requirement, in particular whereby the results of assessment affect grades or educational outcomes (Kasneci et al., 2023; Farrokhnia et al., 2024).

The result aligns with the recent literature that outlines assessment and feedback as high-risk but valuable applications of GenAI. Al-Ali and Miles (2025) have found testing redesign as a crucial field of institutional support needed to facilitate responsible integration. Miao and Holmes (2023) and UNESCO (2024) also highlight the importance of human control, validation, privacy, and accountability in the use of AI in educational evaluation. The current outcome validates these suggestions by demonstrating that a training initiative that is based on reviewing and verifying can be linked to more robust assessment-related practices.

However, its enhancement does not imply a favourable appraisal of autonomous AI. Recent critique warns that even automated feedback can seem personal to be generic, inaccurate, or blind to situational factors of the learner (Farrokhnia et al., 2024; Miao and Holmes, 2023). The most justifiable is the fact that GenAI could be used as a drafting and idea-generation tool, but the responsibility of validity, fairness, accuracy, and final judgment should be left to the teacher.

5.4 Discussion of H04: Overall Instructional Practices

The fourth hypothesis focused on the general GenAI-enhanced instruction practices. The adjusted posttest mean of the experimental group and the control group was 3.98 and 3.09 respectively as reported in Tables 14 and 15. The ANCOVA produced $F(1, 77) = 211.10$, $p < .001$, partial $\eta^2 = .733$. This effect was the most significant in the study and it shows that the intervention was related to the overall enhancement of the lesson planning, instructional design, assessment and feedback.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

The magnitude and consistency of the aggregate outcome are indicative that it was not just a single skill that was affected by the program. Rather, it seems to have underpinned a pattern of professional practice that is integrated. The sessions related immediate design to the curricular alignment and differentiation, verification, assessment validity, privacy, and reflection. Since the same critical cycle was used in various tasks involving teaching, the participants could have acquired a generalizable approach to working with GenAI: stating the educational issue, according to which the contextual information is provided, the output is reviewed and revised, and the professional decision is made regarding the need and manner of its implementation.

Recent models of teacher AI competence uphold this integrated interpretation. In his argument, Celik (2023) proposes that ethical integration and pedagogical integration needs to be interrelated with technological knowledge, pedagogical knowledge, content knowledge, and ethical knowledge. The AI Competency Framework for Teachers of UNESCO also does not distinguish between human-centered judgment, ethics, AI foundations, pedagogy, and professional learning but considers them as inter-related areas instead of distinct technical skills (UNESCO, 2024). The general enhancement observed in the current study fits these frameworks as the intervention factor touched on several dimensions at a time.

The finding aligns with Tan et al. (2025), who reported that a systematic professional development was still required, which related AI knowledge to teaching practice. It also upholds Al-Ali and Miles (2025), who decided that the teacher training, pedagogical support, assessment redesign, and professional networks ought to be built interdependently. Moreover, Chen et al. (2025) concluded that the professional-development requirements of teachers depend on the educational level and encompass not only the elementary AI literacy but also the approaches towards the responsible utilization. The contribution to this literature of the current study is that a coherent program can yield improvement in a number of professional-practice areas in a school situation.

Nevertheless, the overall effect is extremely big and should be interpreted with care. It can be in part a consequence of the high correlation between the content of the training and the objects of the outcome measurement, the promptness of the posttest, the motivation of the participants and the strength of the four-week program. The recent studies are still able to demonstrate that confidence and positive perceptions do not necessarily lead to long-term classroom implementation, and the teachers might still need institutional policies, peer collaboration, technical support, and follow-up learning opportunities (Chen et al., 2025; Li et al., 2026; Tan et al., 2025). Longitudinal measurement and independent study of the lesson plans or classroom performance would be required then to ascertain whether the gains observed last and carry over to normal teaching.

Overall, the four findings suggest that the impact of the program should be viewed as the result of professional learning organized instead of having been exposed to a specific AI tool. The training involved real-life activities, drilling, repetition, clear assessment criteria, feedback and peer review, as well as ethical protection. This interpretation is in line with

ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

the recent findings that the most useful GenAI professional development is when it is customized, context-sensitive and there is an opportunity to test and discuss real instructional products.

6. CONCLUSION

Generative AI is not a threat to education, nor a promise of a solution. It is subject to human capability, pedagogical intent, critical assessment and institutional protection. This research was a quasi-experimental design, which evaluated a structured training intervention using SPSS. The program incorporated timely design and lesson planning, instructional design, lesson differentiation, assessment, verification, privacy and ethics. The reliability estimation, baseline-equivalence test, and correlation test, ANCOVA assumption test, adjusted mean, and effect-size estimation were analyzed. The results indicate that well-organized, pedagogically-based training has the potential to enhance the ability of teachers to incorporate generative AI in planning, instruction, assessment and feedback.

7. FUTURE RESEARCH RECOMMENDATIONS

1) Conduct direct classroom observations to establish whether there are gains in the reported practices by the teachers into real teaching performance.
2) Add follow-up measurements post-training to study the time-sustenance of the effects of the GenAI-based training program.
3) Investigate how GenAI-based professional development affects student learning outcomes, rather than focusing only on teachers’ instructional practices
4) Compare subject-specific training programs with general GenAI training to determine which approach produces greater instructional effects.
5) Research the impact of teacher’s digital competence in the past that determines the effectiveness of GenAI-based teacher development.

References

1) Al-Ali, S., & Miles, R. (2025). Upskilling teachers to use generative artificial intelligence: The TPTP approach for sustainable teacher support and development. *Australasian Journal of Educational Technology, 41*(1), 88-106. https://doi.org/10.14742/ajet.9652
2) Ayanwale, M. A., Sanusi, I. T., Adelana, O. P., Aruleba, K. D., & Oyelere, S. S. (2022). Teachers’ readiness and intention to teach artificial intelligence in schools. *Computers and Education: Artificial Intelligence, 3*, 100099. https://doi.org/10.1016/j.caeai.2022.100099
3) Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence: Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI, 7*(1), 52-62.
4) Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

5) Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. https://doi.org/10.1080/0969595980050102

6) Brandão, A. (2024). Teacher professional development for a future with generative artificial intelligence. *Digital Education Review, 45*, 151-157. https://doi.org/10.1344/der.2024.45.151-157

7) Celik, I. (2023). Towards intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence-based tools into education. *Computers in Human Behavior, 138*, 107468. https://doi.org/10.1016/j.chb.2022.107468

8) Chen, R., Lee, V. R., & Lee, M. G. (2025). A cross-sectional look at teacher reactions, worries, and professional development needs related to generative AI in an urban school district. *Education and Information Technologies, 30*, 16045–16082. https://doi.org/10.1007/s10639-025-13350-w

9) Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 4*, 100118. https://doi.org/10.1016/j.caeai.2022.100118

10) Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating? Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

11) Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education, 20*, 22. https://doi.org/10.1186/s41239-023-00392-8

12) Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute.

13) 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

14) Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., et al. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

15) Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International, 61*(3), 460-474. https://doi.org/10.1080/14703297.2023.2195846

16) Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915-945. https://doi.org/10.3102/00028312038004915

17) Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences, 13*(7), 692. https://doi.org/10.3390/educsci13070692

18) Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching, 8*(3), 381-391. https://doi.org/10.1080/135406002100000512

19) Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

20) Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

21) Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

22) Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? Contemporary Issues in Technology and Teacher Education, 9(1), 60-70.

23) Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.

24) Li, Z., Wang, C., & Bonk, C. J. (2026). Generative AI for teachers' self-directed professional development: A mixed-methods study. *TechTrends, 70*, 23–38. https://doi.org/10.1007/s11528-025-01123-8

25) Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *International Journal of Management Education, 21*(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790

26) Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), 410. https://doi.org/10.3390/educsci13040410

27) Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1-16. https://doi.org/10.1145/3313831.3376727

28) Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO.

29) 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

30) Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2024). Developing language teachers' professional generative AI competence: An intervention study in initial language teacher education. *System, 125*, 103399.

31) Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology, 53*(4), 914-931. https://doi.org/10.1111/bjet.13232

32) Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

33) OECD. (2023). *OECD digital education outlook 2023: Towards an effective digital education ecosystem*. OECD Publishing. https://doi.org/10.1787/c74f03de-en

34) Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14. https://doi.org/10.3102/0013189X015002004

35) Tan, X., Cheng, G., & Ling, M. H. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. *Computers and Education: Artificial Intelligence, 8*, 100355.

36) Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education? *Smart Learning Environments, 10*, 15. https://doi.org/10.1186/s40561-023-00237-x

37) UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO.

38) UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO.

ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

39) UNESCO. (2024). *AI competency framework for teachers*. UNESCO.

40) Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? International Journal of Educational Technology in Higher Education, 16, 39. https://doi.org/10.1186/s41239-019-0171-0

Appendix A: Generative Ai-Enhanced Instructional Practices Scale

Response scale: 1 = strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = strongly agree.

1) I am able to create specific and quantifiable learning goals with the help of generative AI.
2) I am able to generate lesson plans with the help of generative AI in accordance with curriculum outcomes.
3) I am able to organize AI-generated lesson activities in a logical way.
4) I am able to customize AI-generated content according to age and ability levels of students.
5) I am able to recognize prerequisite knowledge of a lesson using generative AI.
6) I am able to come up with different examples of tough concepts.
7) I am able to check the factual correctness of AI-generated lesson materials.
8) I am able to revise the lesson plans created with AI with professional judgment.
9) I am able to create active-learning tasks with the help of generative AI.
10) I am able to produce varied activities to the students with varying needs.
11) I am able to generate numerous variants of explanations of the same idea with the help of generative AI.
12) I am able e to match AI-generated activities with the desired learning outcomes.
13) I am able to apply generative AI in collaborative learning.
14) I am able to recognize the cases of a pedagogically weak activity generated by AI.
15) I am able to adjust AI generated activities to make them more cognitively challenging.
16) I am able to incorporate AI generated materials without diminishing student thinking.
17) I am able to generatively create assessment items based on objectives.
18) I am able to come up with questions that vary in cognitive levels.
19) I am able to generate scoring rubrics with the help of generative AI.
20) I am able to create personalized feedback drafts with the help of generative AI.
21) I am able to detect ambiguity in AI-generated evaluation items.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 59 Issue: 07:2026
DOI: 10.5281/zenodo.21789906

22) I am able to screen AI-generated questions in terms of bias and unsuitable content.
23) I am able to check AI generated answer keys and scoring rationales.
24) I am able to safeguard student privacy in the case of assessment and feedback using generative AI.

Appendix B: Lesson-Plan Performance Rubric

The criteria and four performance levels proposed for evaluating lesson plans are presented in Table B1.

Table B1: Proposed lesson-plan performance rubric

Criterion	1 - Beginning	2 - Developing	3 - Proficient	4 - Advanced
Learning objectives	Objectives are unclear or not measurable	Some objectives are measurable	Objectives are clear and mostly aligned	Objectives are measurable, aligned, and appropriately challenging
Content accuracy	Contains major inaccuracies	Contains minor inaccuracies	Accurate with limited omissions	Accurate, complete, and verified
Instructional coherence	Activities are disconnected	Partial sequence and alignment	Logical sequence and alignment	Coherent, engaging, and strongly aligned
Differentiation	No meaningful differentiation	Superficial adaptations	Appropriate adaptations for some needs	Purposeful and inclusive differentiation
Assessment quality	Assessment does not match objectives	Partial alignment or weak items	Mostly valid and aligned	Strong blueprint, validity, and cognitive range
Feedback design	No usable feedback plan	Generic feedback	Specific feedback for common needs	Actionable, differentiated, and verified feedback
AI-output evaluation	Accepts outputs without review	Limited checking	Identifies and revises weaknesses	Systematic verification, bias review, and revision
Ethics and privacy	Ignores ethical/privacy concerns	Partial awareness	Applies basic safeguards	Transparent, privacy-preserving, and accountable use

As shown in Table B1, the rubric evaluates learning objectives, content accuracy, instructional coherence, differentiation, assessment quality, feedback design, AI-output evaluation, and ethical and privacy safeguards.