AI on AI: Can GenAI Tools Design and Evaluate Course Outlines Better Than We Think?

Jeya Amantha Kumar, Ph.D,
Evidence Driven Learning Innovation (EDLI), College of Arts and Letters (CAL), Michigan State University, East Lansing, Michigan, USA (davidpan@msu.edu / jeya.amantha@gmail.com)

Abstract
Despite the increasing use of generative AI (GenAI) tools in education, little is known about their effectiveness in producing pedagogically sound instructional materials. Therefore, this study evaluated the performance of six GenAI tools as instructional designers in generating a unit or module outline for an undergraduate course, focusing on developing learning objectives based on Universal Design for Learning (UDL) principles and later evaluating each outcome. Six free versions of GenAI, ChatGPT, Claude, Copilot, Gemini, Meta AI, and Perplexity were then analyzed thematically, focusing on instructional strategies, UDL integration, accessibility, and rubric development, while also being evaluated using a standardized points-based rubric by each GenAI tool. Findings revealed that Perplexity, Claude, and Gemini consistently produced stronger, learner-centered outlines, while Copilot and Meta demonstrated weaker instructional coherence. ChatGPT demonstrated strong instructional coherence but showed limitations in depth of rubric and integration of accessibility. Additionally, common limitations included insufficient timeline structuring, limited integration of learning domains beyond cognition, and weak alignment to summative assessments.

Keywords: Generative AI in Education; Instructional Design; Universal Design for Learning; AI-Based Instructional Evaluation; AI-Enhanced Teaching

1. Introduction

The rise of generative artificial intelligence (GenAI) tools has prompted significant interest in their integration within higher education, particularly among educators exploring innovations in curriculum design (Abbasi et al., 2025; Dennis, 2025), assessment practices, and co-development of instructions (Kostikova et al., 2024; Lang et al., 2025). Consequently, large language models (LLMs), especially Generative Pre-trained Transformers (GPTs) such as ChatGPT (OpenAI, 2025) and Copilot (Microsoft, 2025), have been widely adopted not only to support such educational transformation but also to enhance learning engagement (Abbasi et al., 2025). However, as rapidly as GenAI can generate content, scholars have cautioned against overreliance (Parker et al., 2024), as it has been highlighted as a potential threat to undermine academic integrity and quality. Key risks include the potential misalignment between AI-generated outputs and intended learning outcomes, misinformation (Qadir, 2023), hallucinated content (Davis & Lee, 2024), and the reinforcement of existing biases (Khamis et al., 2025; Kostikova et al., 2024). However, a key drawback in these findings is that most education-based research centers mainly on ChatGPT by OpenAI (Lang et al., 2025). A simple Scopus database search (Jun 2025) combining “AI” and “Education” with specific GenAI tools showed a dominant focus on ChatGPT (N=4,285 publications since 2022), compared to significantly fewer for Gemini AI (N=9 since 2024), Microsoft Copilot (19, since 2022), Claude AI (N=9, since 2023), Perplexity AI (N=15, since 2023), and Meta AI (19, since 2023), underscoring ChatGPT’s overwhelming presence in educational research. Al-khresheh (2024) highlighted that this could lead to unexplored biases when the effectiveness and consistency of GenAI in instructional design are judged solely based on ChatGPT’s effectiveness.

Similarly, this trend is frequently observed in studies related to course development, equally highlighting further investigation to support educators’ adoption (Kumar et al., 2024). Nevertheless, developing a well-designed curriculum requires more than content generation, as it involves the careful and complex mapping between learning objectives, instructional strategies, assessments, rubrics (Ghonoodi & Salimib, 2011), course outlines, and outcomes to ensure pedagogical coherence (Chan & Colloton, 2024; Ullmann et al., 2024), all while integrating Universal Design for Learning (UDL) principles (Kumar et al., 2024) and alignment with instructional frameworks such as Bloom’s Taxonomy (Dickey & Bejarano, 2023; Karpouzis et al., 2024). While GenAI tools are capable of generating instructional content and evaluating such outputs using standardized criteria, this study examines how six distinct GenAI models perform in this role when prompted to develop course unit outlines and further evaluate their own and peer-generated content using a consistent prompt and rubric. The six GenAI selected are ChatGPT, Gemini, Copilot, Claude, Perplexity, and Meta AI, and the following research questions guided this study:

RQ1: Which GenAI tool produces the highest-quality learning unit outlines, as measured by a standardized instructional design rubric across dimensions such as clarity and measurability of learning objectives (LO), alignment between LO, instruction, and assessment, UDL integration, accessibility, and overall instructional quality and coherence as measured by GenAI tools?

RQ2: How do these tools vary in designing and evaluating these instructional units?

2. Methodology

2.1 AI Tools as respondents

ChatGPT, Gemini (Google, 2025), Copilot, Claude (Anthropic, 2025), Perplexity AI (Perplexity AI, 2025), and Meta AI (Meta AI, 2025) were engaged as non-human participants across two phases: content developers in Phase 1 and reviewers in Phase 2. All tools were accessed through their free versions without login, except for Claude, which required user authentication to enable free access. Additionally, to maintain consistency in outputs, two standardized prompts were created using the ChatGPT 4o model (Appendix A).

2.2 Prompting Procedure

Phase 1: Unit/Module outline creation

Each GenAI were prompted to develop a course outline that consists of a unit or module title, a brief overview consisting of two to three sentences, two to three well-defined learning objectives, a concise outline of instructional activities, one to two formative assessment ideas, and a bulleted rubric to evaluate student performance. The prompt also instructed each AI to avoid table formatting and present rubrics as bulleted lists to promote consistency. All outlines were expected to reflect UDL principles and align with pedagogical requirements in U.S. higher education.

Phase 2: Peer Evaluation using rubric-based assessment

The AI-generated unit/module outlines were next evaluated through a reciprocal review process using a structured, 20-point rubric. Each AI tool reviewed the outlines produced by the other five based on five equally weighted instructional design criteria: clarity and measurability of learning objectives, alignment between objectives, instructional activities, assessment methods, integration of UDL principles, accessibility, and overall instructional quality and coherence.

2.3 Procedure and Analysis

The GenAI models used were ChatGPT-GPT-4-turbo, Gemini 2.5 Flash, Copilot, Claude 3.7 Sonnet, Perplexity AI (vGPT-4o), and Meta AI-Llama 4. No identifiable information related to the GenAI was included in any course outline when uploading different outlines. Each model's course outline was then thematically examined to reveal patterns in how various AI tools interpret and implement instructional design principles, with all quantitative analyses conducted using Microsoft Excel.

3. Findings and Discussion

Evaluations of AI-generated output (Table 1) revealed that Perplexity achieved the highest average score (M = 97.33), followed closely by Gemini (M = 96.33) and Claude (M = 96.00). Copilot (M = 84.83) and Meta (M = 80.50) received lower overall scores, while ChatGPT was considered average (M = 93.50). A comparative analysis of each reviewer's highest and lowest scoring units substantiates these findings, with Perplexity consistently recognized as a top performer, as evaluated by Gemini, Meta, and Perplexity itself. In contrast, Meta AI obtained the lowest scores according to all GPT reviewers, except for Meta itself. Additionally, Copilot was also noted among the lowest performers, as assessed by Meta and Perplexity.

Table 1: *Comparative Evaluation of the Six GenAI Tools*

GenAI	Highest		Lowest		Clarity & Measurability (20)	Alignment (20)	UDL (20)	Accessibility (20)	Instructional Quality (20)	Total
	GenAI	Score	GenAI	Score						
ChatGPT	ChatGPT	100	Meta	82	18.83	19.17	18.50	17.83	19.00	93.50
Claude	Gemini	99	Meta	71	19.50	19.50	19.17	18.33	19.50	96.00
Co-Pilot	Claude	93	Meta	83	14.67	17.33	16.33	15.83	17.33	84.83
Gemini	Perplexity	99	Meta	70	19.50	19.17	15.58	18.83	19.50	96.33
Meta	Perplexity	100	Copilot	92	16.00	16.50	15.50	15.17	16.50	80.50
Perplexity	Perplexity	100	Copilot & Meta	83	19.67	19.67	19.83	19.50	19.83	97.33

Next, the analysis of unit/module titles (Appendix B) revealed a focus on developing LO, with ChatGPT, Meta, and Perplexity emphasizing skills development through action-oriented phrases such as "writing" and "crafting." In contrast, Claude, Copilot, and Gemini framed it within a broader pedagogical context, with Claude highlighting the shift "from theory to practice," while Copilot and Gemini concentrated on associating LO with foundational instructional design principles. Furthermore, it was noted that each set of LOs reflected different cognitive levels of Bloom's Taxonomy, where Copilot and Gemini included lower-order skills (e.g., define, identify), Chatgpt and Meta emphasized mid- to high-level skills (apply, evaluate), and Claude and Perplexity targeted higher-order thinking (analyze, develop, and evaluate). Notably, Gemini, Meta, and Perplexity developed their objectives using specific instructional design frameworks, with Gemini focusing on the Audience, Behavior, Condition, and Degree (ABCD) model and Meta and Perplexity structuring their LO around UDL. Although all tools produced three relevant objectives, the use of higher-order Bloom's levels appears to influence peer-reviewed outcomes, as Perplexity (M = 19.67), Gemini (M = 19.5), and Claude (M = 19.5) received the highest average scores for Clarity and Measurability. In contrast, Copilot (M = 14.67) and Meta (M = 16.0), which relied more on lower- to mid-level skills, received moderately lower ratings.

Following this, based on the suggested instructional activities, ChatGPT and Copilot incorporated traditional, in-person strategies such as lectures and case studies, supplemented by group work, peer review, and hands-on practice. Gemini and Perplexity emphasized highly interactive and learner-centered designs, utilizing strategies like polls, think-pair-share, role-playing, and scenario analysis to foster engagement and collaboration. However, Claude was the only model that presented a detailed, scaffolded three-day instructional plan integrating lectures, multimedia elements, collaborative tasks, and feedback opportunities. In contrast, while proposing a more conventional structure centred on lectures and guided practice, Meta offered fewer modalities. Overall, Claude, Gemini, and Perplexity showed greater pedagogical richness, with Perplexity adding resource links and follow-up questions.

Regarding formative assessment strategies, all six GenAI tools consistently included similar self-assessment tools, such as self-check quizzes and peer reviews. Although these activities were presented as suggestions in the prompt, ChatGPT was the only platform offering a distinctive recommendation to integrate a reflection journal. However, despite this, rubrics did vary, with ChatGPT, Claude, and Gemini aligning their evaluation criteria with instructional design principles such as clarity, measurability, alignment, and UDL integration, though they differed in scale structure. Claude used a broader 0–25 range, emphasizing content knowledge and structural quality, while Gemini integrated ABCD model elements into structure evaluation. In contrast, Copilot applied a vague 5-point Likert scale with no clear criteria, Perplexity limited its rubric to peer review, and Meta's 8-point model offered scoring consistency but lacked evaluative depth. Overall, ChatGPT, Claude, and Gemini demonstrated more robust and instructionally aligned rubric development, whereas Copilot, Meta, and Perplexity exhibited reduced comprehensiveness and rigor. However, when compared to the criteria focusing on alignment between objectives, instruction, and assessment, Perplexity (M = 19.67), Claude (M = 19.5), ChatGPT, and Gemini (both M = 19.17) achieved the highest average scores (as evaluated by other GPTs), indicating strong instructional coherence. In contrast, Copilot (M = 17.33) and Meta (M = 16.5) demonstrated weaker alignment, reflecting instructional flow and integration gaps. Perplexity scored highly overall despite its limited rubric because its strong alignment was in one area, which was peer review, while others focused on a more general evaluation of skills in developing LO. However, what was observed was that while there was alignment, the rubric may have lacked applicability with other assessments, highlighting GPT's capability as an evaluator.

Next, for UDL integration, Perplexity (M = 19.83), Claude (M = 19.17), and ChatGPT (M = 18.5) demonstrated the strongest applications. Perplexity stood out for its high flexibility in assessment forms, use of editable templates, audio guides, and integration of real-world role-play, offering learners multiple means of expression and engagement. Claude incorporated multimedia tools, peer feedback, and scaffolded activities that supported representation and action, whereas ChatGPT emphasized multimodal engagement through lectures, group work, case studies, and opportunities for self-reflection. In contrast, Copilot (M = 16.33) focused more narrowly on objective refinement and accessibility, with limited support for diverse learning formats. Gemini (M = 15.58), although blending UDL principles with the ABCD model, showed weaker integration across the full range of UDL checkpoints, whereas Meta (M = 15.5) applied UDL more restrictively, focusing primarily on revising and writing tasks for diverse learners without broader multimodal or adaptive engagement strategies. Lastly, in terms of accessibility features, Perplexity (M = 19.5) and Gemini (M = 18.83) demonstrated the most comprehensive approaches. Perplexity offered editable templates, audio guides, and a feedback loop, enabling flexible learner interaction and customization. Gemini, on the other hand, supported multimodal access through audio, video, infographics, and text-to-speech tools, alongside attention to pacing and varied content formats. Claude (M = 18.33) followed closely, providing visual, audio, and text-based formats with scaffolding and learner choice, though slightly less emphasis was placed on assistive technology. ChatGPT (M = 17.83) provided strong accessibility through written content, video with captions, and audio, though it leaned more heavily on content delivery rather than adaptive features. In contrast, Copilot (M = 15.83) acknowledged accessibility in general terms but lacked detailed implementation, while Meta (M = 15.17) concentrated its accessibility support on inclusive writing practices and peer feedback examples, with limited modality variation.

Finally, when considering overall instructional quality and coherence, Perplexity (M = 19.83), Claude, Gemini (both M = 19.50), and ChatGPT (M = 19.00) received the highest ratings, reflecting strong pedagogical structure, clarity, and instructional flow in their unit or module designs. ChatGPT, Claude, and Copilot shared common strengths in applying established instructional frameworks (i.e., Bloom's Taxonomy, SMART) while also providing opportunities for peer collaboration and formative feedback. Claude, in particular, stood out for its structured sequencing from theory to practice, multimodal instructional design, and strong scaffolding strategies. Likewise, Gemini and Perplexity emphasized real-world application, mastery learning, and technology integration, aligning their instructional strategies with authentic practice and learner flexibility. Despite these strengths, a common shortfall across tools was the absence of clearly defined timelines for instructional activities, which may hinder pacing and learner expectations. ChatGPT, Copilot, and Meta also exhibited limitations in rubric design, lacking depth or omitting full scoring ranges (i.e., ChatGPT). While only one rubric was created for all formative assessments (none for self-check assessment), Perplexity's high score could be due to how its rubric aligned with the formative assessment; however, the more general rubric by other GenAI were found to focus on the LO itself, making it more applicable for numerous activities. Nevertheless, only Claude attempted to associate activities with summative assessment. Additionally, none of the GenAI tools assessed learning outcomes beyond the cognitive domain, such as the affective and psychomotor domains. Furthermore, based on user experience, it was observed that all the free versions (without login) lack the capability for users to upload documents. However, Perplexity stood out, allowing users to effortlessly download documents in either PDF, DOC, or Markdown, presented in a clean and organized fashion, and this feature can be particularly advantageous for instructors seeking to streamline their resources. Moreover, Perplexity takes it a step further by offering insightful suggestions for follow-up questions and indicating the sources of the information provided, thereby creating transparency and trust in the content. On the other hand, while Meta did not perform as well as the others, it took an unusual approach of requiring users to specify their age, aiming to tailor feedback to their demographic. This raises an intriguing question: might the generated content vary according to the age groups of its users? The response remains uncertain, prompting curiosity about the potential influence of age on the responsiveness of these platforms.

4. Conclusion and Future Directions

The findings demonstrated different levels of instructional alignment, cognitive skills integration, UDL adherence, and accessibility. Perplexity, Claude, and Gemini consistently stood out for their structured, learner-centered designs and flexible engagement strategies, while ChatGPT also demonstrated strong instructional coherence but showed minor limitations in rubric depth and accessibility integration. In contrast, Copilot and Meta exhibited comparatively weaker instructional coherence and rubric development. Limitations were also observed in the absence of clear timelines, the lack of consideration for affective and psychomotor domains, and the alignment with summative assessments. Platforms prioritising clarity, accessibility, and real-world applicability received stronger evaluations from other GenAI. These findings show that while GenAI tools can design and evaluate instructional units reasonably well, they still have significant gaps in activity structuring, holistic domain integration, and alignment with summative assessments. Furthermore, inter-rater reliability was not assessed, as GenAI tools differ in architecture and heuristics, where their divergence likely reflects model design rather than inconsistent interpretation of the rubric. Future studies could include Grok (xAI, 2025); although not formally analyzed, preliminary testing showed Grok offering an engaging user experience through emojis, humor, and wit, while objectively rating Perplexity the highest, followed by Claude, and assigning itself a score of 89.

Reference

Abbasi, B. N., Wu, Y., & Luo, Z. (2025). Exploring the impact of artificial intelligence on curriculum development in global higher education institutions. *Education and Information Technologies*, *30*(1), 547–581. https://doi.org/10.1007/s10639-024-13113-z

Al-khresheh, M. H. (2024). Bridging technology and pedagogy from a global lens: Teachers' perspectives on integrating ChatGPT in English language teaching. *Computers and Education: Artificial Intelligence*, *6*(December 2023), 100218. https://doi.org/10.1016/j.caeai.2024.100218

Anthropic (2025). Claude 3.7 Sonnet [AI assistant software]. https://www.anthropic.com. Retrieved April 29, 2021.

Chan, C. K. Y., & Colloton, T. (2024). *Generative AI in Higher Education : The ChatGPT Effect*. Routledge. https://doi.org/10.4324/9781003459026

Davis, R. O., & Lee, Y. J. (2024). Prompt: ChatGPT, Create My Course, Please! *Education Sciences*, *14*(1), 1–12. https://doi.org/10.3390/educsci14010024

Dennis, M. J. (2025). How will AI influence higher education? *Enrollment Management Report*, *28*(12), 3–8. https://doi.org/10.1002/emt.31358

Dickey, E., & Bejarano, A. (2023). A Model for Integrating Generative AI into Course Content Development. *A Model for Integrating Generative AI into Course Content Development*. https://doi.org/https://doi.org/10.48550/arXiv.2308.12276

Ghonoodi, A., & Salimib, L. (2011). The study of elements of curriculum in smart schools. *Procedia - Social and Behavioral Sciences*, *28*, 68–71. https://doi.org/10.1016/j.sbspro.2011.11.014

Google (2025). Gemini [Large language model]. https://gemini.google.com/ . Retrieved April 21, 2025.

Karpouzis, K., Pantazatos, D., Taouki, J., & Meli, K. (2024). Tailoring Education with GenAI: A New Horizon in Lesson Planning. *2024 IEEE Global Engineering Education Conference (EDUCON)*, 1–10. https://doi.org/10.1109/EDUCON60312.2024.10578690.12071

Khamis, N., Chen, B., Egan, C., Gaglani, S., & Tackett, S. (2025). More intelligent faculty development: Integrating GenAI in curriculum development programs. *Medical Teacher*, *0*(0), 1–3. https://doi.org/10.1080/0142159X.2025.2473606

Kostikova, I., Holubnycha, L., Besarab, T., Moshynska, O., Moroz, T., & Shamaieva, I. (2024). Chat GPT for Professional English Course Development. *International Journal of Interactive Mobile Technologies (IJIM)*, *18*(02), 68–81. https://doi.org/10.3991/ijim.v18i02.46623

Kumar, J. A., Zhuang, M., & Thomas, S. (2024). ChatGPT for natural sciences course design : Insights from a strengths , weaknesses , opportunities , and threats analysis. Natural Science Education, November, 1–19. https://doi.org/10.1002/nse2.70003

Lang, Q., Wang, M., Yin, M., Liang, S., & Song, W. (2025). Transforming Education with Generative AI (GAI): Key Insights and Future Prospects. IEEE Transactions on Learning Technologies, 18, 230–242. https://doi.org/10.1109/TLT.2025.3537618

Meta AI (2025). Llama 4 [Large language model]. https://www.meta.ai/ . Retrieved April 21, 2025.

Microsoft (2025). Microsoft Copilot [Large language model]. https://copilot.microsoft.com . Retrieved April 21, 2025.

OpenAI (2025). ChatGPT [Large language model]. https://chat.openai.com/ . Retrieved April 21, 2025.

Parker, L., Carter, C., Karakas, A., Loper, A. J., & Sokkar, A. (2024). Graduate instructors navigating the AI frontier: The role of ChatGPT in higher education. Computers and Education Open, 6(February), 100166. https://doi.org/10.1016/j.caeo.2024.100166

Perplexity AI (2025). Perplexity [Large language model]. https://www.perplexity.ai. Retrieved April 21, 2025.

Qadir, J. (2023). Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education. IEEE Global Engineering Education Conference, EDUCON, 2023-May. https://doi.org/10.1109/EDUCON54358.2023.10125121

Ullmann, T. D., Edwards, C., Bektik, D., Herodotou, C., & Whitelock, D. (2024). Towards Generative AI for Course Content Production: Expert Reflections. European Journal of Open, Distance and E-Learning, 26(s1), 20–34. https://doi.org/10.2478/eurodl-2024-0013

xAI (2025). Grok [Large language model]. https://grok.com/ . Retrieved June 26, 2025.

Appendix A

Prompts for Phase 1: Unit/Module Outline Creation Prompt
You are a higher education instructor teaching undergraduate students in education how to design and develop instruction.
Please create a unit/module outline to help these students learn how to write effective learning objectives. This unit should represent one part of a broader instructional design course.
Your outline must be guided by Universal Design for Learning (UDL) principles and reflect expectations in U.S. higher education.

Please include the following components:
- A clear and relevant unit/module title
- A brief overview of the unit/module (2–3 sentences)
- 2–3 learning objectives that are measurable and clearly written
- A short outline of the instructional activities or teaching strategies planned
- 1–2 formative assessment ideas (e.g., reflection, peer review, self-check quiz)
- A rubric to evaluate student performance on the assessment

◆ Important formatting instruction:
- Avoid using tables
- Present the rubric as a bulleted list with clear point allocation for each level of performance

Use section headings and focus on instructional alignment, UDL integration, and accessibility.

Prompts for Phase 2: Peer Evaluation Prompt (with Detailed Rubric)

Instruction:
Generate a structured comparison of six unit/module outlines designed to teach undergraduate education students how to write effective learning objectives. Use the provided criteria to evaluate and score each outline, then present the results in a comparative table with key highlights for each criterion. Please ask me for the 6 units/ modules.

Rubric for Evaluation (Total: 100 Points)
1. Clarity and Measurability of Learning Objectives (20 pts)
- **18–20**: Objectives are clear, specific, and measurable, using action verbs and aligned with Bloom's Taxonomy.
- **14–17**: Mostly clear and measurable, with some minor vagueness.
- **10–13**: Objectives are present but vague or hard to measure.
- **0–9**: Objectives are too broad, missing, or not measurable.

2. Alignment Between Objectives, Instruction, and Assessment (20 pts)
- **18–20**: Instruction and assessment clearly support and align with all objectives.
- **14–17**: Mostly aligned with some minor disconnects.
- **10–13**: Partial alignment; some gaps between objectives and assessments.
- **0–9**: Major misalignment or unclear connections between components.

3. Integration of UDL Principles (20 pts)
- **18–20**: Multiple UDL principles (e.g., engagement, representation, expression) are clearly applied.
- **14–17**: Some UDL elements present, but not consistently integrated.
- **10–13**: Minimal UDL integration; mostly implicit or underdeveloped.
- **0–9**: No evidence of UDL consideration.

4. Accessibility (20 pts)
- **18–20:** Instruction is accessible and responsive to diverse learner needs (e.g., accommodates different learning preferences, includes alt text, and offers flexible content formats).
- **14–17:** Demonstrates general attention to accessibility and adaptation for varied learners.
- **10–13:** Shows limited consideration for learner diversity or accessibility features.
- **0–9:** Lacks evidence of accessibility features or adaptation for different learner needs.

5. Overall Instructional Quality and Coherence (20 pts)
- **18–20**: Content is organized, coherent, and pedagogically sound.
- **14–17**: Mostly clear and effective, with minor inconsistencies.
- **10–13**: Some good ideas, but lacks flow or consistency.
- **0–9:** Poor structure, unclear intent, or fragmented content.

Appendix B

Criteria	ChatGPT	Claude	Co-pilot	Gemini	Meta	Perplexity
Title	Writing Effective Learning Objectives for Instructional Design	Creating Effective Learning Objectives: From Theory to Practice	Crafting Effective Learning Objectives: The Foundation of Instructional Design	Foundations of Instructional Design: Writing Measurable Learning Objectives	Crafting Effective Learning Objectives	Crafting Measurable Learning Objectives in Instructional Design
Learning objectives	Write, Evaluate, Apply	Analyze, Develop, Evaluate -	Define, Apply, Evaluate	Identify, Evaluate, Write with ABCD Model	Write, Apply, Evaluate/Revise with UDL	Distinguish, Construct, Design with UDL
Instructional activities	Lecture, group SMART analysis, workshop, peer review, case study	3-day plan: lectures, group, multimedia, card sort, lesson planning, feedback	Lecture, case study, hands-on application with feedback	Poll, lecture, videos, templates, scenarios, think-pair-share	Lecture, guided practice, case study, online resources	Case analysis, multimodal resources, role-play, templates
Formative assessments	Self-check quiz, reflection journal	Workshop, self-assessment with feedback	Peer review workshop, self-check quiz	Objective analysis check, peer-reviewed draft	Peer review, self-check quiz	Self-check quiz, peer review exchange
Rubric	4-Point Likert Scale (2–5): Criteria include Clarity & Measurability, Alignment, UDL, Writing Quality.	4-Level Scale (0–25) Criteria: Content Knowledge, Structure and Format, Alignment and Application, UDL Considerations	5-Point Likert Scale (0–4): Levels include Not Attempted (0), Needs Improvement (1), Developing (2), Proficient (3), Exemplary (4). No criteria	5-Point Likert Scale (1–5): Criteria - Clarity, Measurability, ABCD Components, Alignment.	4-Level Scale (1–8 total): Novice (1–2), Developing (3–4), Proficient (5–6), Advanced (7–8), 2 points per each of 4 criteria.	4-Point Likert Scale (1–4): Levels are Incomplete (1), Developing (2), Proficient (3), Exemplary (4); No criteria and only focus on peer review
UDL integration	Multiple formats and flexible assessments: lecture, group, case, self-reflection	Multimedia, peer feedback, flexible demo options	Objective refinement and accessibility integration	ABCD + UDL combo, multimodal representation, engagement	UDL in revising and writing for diverse learners	Strong UDL emphasis in flexibility and assessment forms
Accessibility features	Written, video with captions, and audio recordings	Visual/audio/text formats, pacing, choice, and scaffolded content	Accessible formats, varied learning needs support	Audio, video, reading, templates, infographic, and TTS support	Inclusive writing focus, peer feedback, accessible examples	Editable templates, audio guides, and a feedback loop.