AI in the Classroom

A rapid evaluation of AI-generated lesson plans in Islamabad's public schools

UK International Development — Partnership | Progress | Prosperity

THE WORLD BANK

unicef for every child

About this document

Reviewers

Molly Jamieson Eberhardt, Joel Mitchell, Laila Friese

About EdTech Hub

EdTech Hub is a global research partnership. Our goal is to empower people by giving them the evidence they need to make decisions about technology in education. Our evidence library is a repository of our latest research, findings, and wider literature on EdTech. As a global partnership, we seek to make our evidence available and accessible to those who are looking for EdTech solutions worldwide.

EdTech Hub is supported by UKAid, Bill & Melinda Gates Foundation, World Bank, and UNICEF. The views in this document do not necessarily reflect the views of these organisations.

To find out more about us, go to edtechhub.org/. Our evidence library can be found at docs.edtechhub.org/lib/.

Contents

Figures

Tables

Abbreviations and acronyms

GK   General knowledge

LP   Lesson plan

NIETE   National Institute of Excellence in Teacher Education

AI   Artificial Intelligence

LMIC   Low- and middle-income country

1. Introduction

The use of AI in education is transforming how instructional content is created for, delivered, and adapted to diverse learning environments. AI-powered tools can streamline time-intensive tasks such as lesson planning, assessment generation, and grading, allowing teachers to focus more on student engagement and individualised instruction. In low- and middle-income countries (LMICs) like Pakistan, where teacher capacity and resource limitations often pose significant challenges, AI-generated materials present an opportunity to standardise content quality while reducing the planning burden on educators.

This report evaluates AI-generated lesson plans developed by Taleemabad, an educational technology company in Pakistan, within the framework of the GSMA Innovation Fund for Accelerated Growth and through their implementation in the National Institute of Excellence in Teacher Education (NIETE) programme. Initiated in January 2024, the NIETE programme is a partnership between the Federal Directorate of Education and Taleemabad. The expansive programme, begun under the auspices of the Ministry of Federal Education and Professional Training (MoFEPT), currently operates across 341 public schools in Islamabad Capital Territory, implementing what is described as a "combo model" that incorporates digital learning tools with in-person mentoring. The NIETE programme consists of three main elements:

1. A digital continuous professional development programme for teachers

2. The generation and use of standardised, AI-generated lesson plans aligned with the national curriculum

3. Provision of coaching for teachers through classroom observations using the TEACH tool.

The programme currently encompasses approximately 4,044 primary school teachers and 98,000 students throughout six regions in the Islamabad Capital Territory. Drawing influence from international approaches such as Singapore's "graduates as multipliers" framework, NIETE aims to develop a network of teacher-leaders who can support educational improvement within their local contexts.

Conducted by EdTech Hub, this evaluation leverages expertise in educational technology research to deliver a nuanced understanding of how the centrally developed AI-generated lesson plans can enhance

equitable access to high-quality learning content. By triangulating data from diverse methodologies, the findings aim to guide strategic improvements in Taleemabad's AI-powered lesson plans and NIETE's deployment of the lesson plans, fostering innovation while addressing systemic educational challenges in LMICs.

1.1. Potential impact of this study

By evaluating the quality and usability of AI-generated lesson plans within the NIETE programme,[1] this study provides valuable evidence to guide the effective integration of AI in public education. The insights from this research are especially timely as the programme is being expanded to other provinces in Pakistan, starting with Balochistan. The findings will play a critical role in shaping the implementation strategy by identifying best practices and potential areas for improvement, ensuring that the AI-generated lesson plans can be adapted to diverse educational contexts. This evidence will support policymakers, education leaders, and technology developers in making informed decisions as they scale AI-driven educational initiatives across the country, fostering more consistent and effective teaching practices nationwide.

1.2. Evaluation focus

The evaluation centres on assessing the quality and efficacy of Taleemabad's centralised AI-generated lesson plans in improving access to high-quality learning content across diverse educational contexts. Key areas of focus include the alignment of AI-generated lesson plans with curricular standards, their adaptability to the needs of students and teachers, and their influence on pedagogical practices. Additionally, the evaluation examines stakeholder perceptions to identify areas for improvement, ensuring that the tools are both contextually relevant and scalable for broader adoption.

The absence of baseline data limits our ability to measure shifts in instructional effectiveness over time. Instead, our analysis focuses on evaluating the quality, relevance, and adaptability of the lesson plans based on teacher perceptions and expert assessments. Additionally, since teachers were unaware that these lesson plans were AI-generated, the study does not assess implementation dynamics related to teacher attitudes toward AI-driven content creation. The scope of this study is also

---

[1] In this report, we refer to the AI-generated lesson plans as such, or as Taleemabad's lesson plans or NIETE's lesson plans, using these terms interchangeably depending on context.

limited to the lesson plans' design and usability, rather than their immediate impact on student learning outcomes.

1.2.1. Evaluation questions

The evaluation questions used for this study were:

Quality of AI-generated lesson plans

How do NIETE's AI-generated lesson plans compare to traditional teacher-created plans in terms of relevance, content quality, and adaptability to diverse contexts and learner needs?

Impact on pedagogical practices

To what extent have AI-generated lesson plans influenced teaching methodologies and classroom practices among educators in urban and rural settings?

Perception and usability

What are the teachers' perceptions of NIETE's centrally developed lesson plans,[2] and which aspects are identified as strengths or areas needing improvement?

---

[2] Although the lesson plans were AI-generated by Taleemabad, teachers were not aware of this, so they did not associate them directly with Taleemabad or AI.

2. Literature review

As AI becomes increasingly integrated into education, a growing area with potential lies not just in automating lesson planning, but in equipping teachers with high-quality, adaptable content that enhances learning outcomes across diverse contexts. In order to realise that potential, it is imperative to consider pedagogical depth, local relevance, and potential limitations in content design and development.

To contextualise our research project, we reviewed existing literature on the increasingly innovative role of AI-powered educational tools in improving content quality. Our review also examines global and regional applications of AI-generated lesson plans, the impact of centralised versus decentralised repositories, and urban-rural accessibility. Additionally, we explored the broader debates surrounding AI's role in education, such as balancing efficiency with pedagogical depth, automation with teacher agency, and inclusion with bias mitigation.

There is evidence to suggest that artificial intelligence (AI) is transforming education by enhancing learning, streamlining administrative tasks, and personalising content (⬆Wang et al., 2024). Key applications include adaptive learning platforms that adjust to student needs, intelligent tutoring systems providing real-time feedback, and automated grading systems reducing educators' workloads. Adaptive learning uses AI to tailor instruction, improving accessibility and efficiency (⬆Strielkowski et al., 2024). Intelligent tutoring systems enhance engagement by offering personalised guidance (⬆Xu, 2024). Automated grading saves time, allowing educators to focus on teaching (⬆Wang et al., 2024). However, the adoption of AI in education presents a key challenge of how to ensure that tools mitigate rather than augment preexisting inequalities and weaknesses in educational practices (⬆UNESCO, 2024b). Additionally, policymakers must develop frameworks that align AI use with broader educational goals, supporting localised, context-aware implementations.

According to two recent studies by ⬆Fan et al. (2024) and ⬆Karpouzis et al. (2024), teachers using AI tools in China and Greece respectively experienced a significant decrease in workload and an improvement in the quality of their lesson plans, suggesting that AI tools can effectively support educators in instructional design. AI-powered lesson planning tools like MagicSchool.ai and Diffit have also gained traction, reportedly reducing teacher workload by 31% while maintaining lesson quality (⬆Education Endowment Foundation, 2024). However, while AI can enhance educational efficiency, its deployment must be accompanied by

empirical research to validate its effectiveness across diverse learning environments (↑Gurl et al., 2024). It should also ensure that it augments rather than hampers the use of critical skills by educators and students (↑Bastani et al., 2024). Such considerations highlight the need for structured evaluation frameworks that measure AI's pedagogical impact beyond theoretical benefits.

AI-powered tools are also being successfully leveraged to enhance pedagogical content quality by automating the generation of lesson plans, quizzes, and instructional materials aligned with curriculum standards(↑Baytak, 2024). There is research to indicate that high-quality lesson materials improve student performance, particularly in under-resourced settings where teachers may lack formal training (↑Education Endowment Foundation, 2024). AI-generated lesson plans can ensure curriculum alignment and reduce variability in teaching quality. However, studies highlight the need for AI tools to support differentiated instruction, as AI-generated lesson plans often lack customisation for diverse learners (↑Baytak, 2024). Structured evaluation methodologies must be used to ensure the quality and effectiveness of AI-generated lesson plans. In addition to established frameworks such as Bloom's Taxonomy (↑Anderson & Krathwohl, 2001), Universal Design for Learning (UDL) (↑Meyer et al., 2014), and Backward Design (↑Wiggins & McTighe, 2005), there is growing advocacy for teacher-led evaluation protocols. ↑Gurl et al. (2024) emphasise that teachers must critically assess AI-generated content, ensuring it aligns with student needs rather than passively adopting AI-driven outputs.

Research on centralised repositories such as the UK's AI content bank suggests that while these offer vetted, high-quality content, decentralised platforms like TeachersPayTeachers[3] allow for greater teacher customisation but risk inconsistencies (↑GOV.UK, 2024; ↑Harris et al., 2023). The ideal model may involve a hybrid approach, where structured AI-generated content is paired with flexible teacher-driven modifications (↑Molina et al., 2024).

Global examples illustrate AI's growing potential. In Chile, UmmIA,[4] an AI-powered lesson planning tool that offers structured but adaptable content, is being rapidly implemented. It has been designed to support teachers without overstepping their professional expertise and judgement. Similarly, in the United States, MagicSchool.ai,[5] created by former teachers,

---

[3] See https://www.teacherspayteachers.com/. Retrieved 27 March 2025.

[4] See https://ummia.cl/. Retrieved 27 March 2025.

[5] See https://www.magicschool.ai/. Retrieved 27 March 2025.

helps educators generate lesson plans, assessments, and AI-resistant assignments, streamlining their workload while maintaining lesson quality. In India, Microsoft's Shiksha Copilot[6] and MagicSchool.ai are being employed to provide government school teachers with AI-assisted lesson plans aligned with Central Board of Secondary Education curricula (↑Molina et al., 2024). A field test in Karnataka found that 90% of teachers reduced lesson planning time from over an hour to just 5–15 minutes, allowing them time to focus more on classroom engagement and student interaction (↑Uygun, 2024). Despite these successes, ensuring local and cultural relevance remains a challenge, as AI-generated examples may not always align with students' lived experiences (↑OECD, 2024). The debate between structured AI-generated content and teacher autonomy is significant, with studies showing that overly prescriptive AI lesson plans can limit teachers' ability to contextualise material effectively (↑Akeyampong et al., 2023; Frøsig & Romero, 2024).

In Bangladesh, the ↑Kong et al. (2024) suggests AI could assist teachers with curriculum-aligned planning, provided localised adaptations are prioritised. In Southeast Asia, while Indonesia and Malaysia are exploring AI-powered content creation, Singapore has taken a policy-driven approach, integrating AI-generated instructional materials at the national level (↑Kong et al., 2024; ↑Ministry of Education, Singapore, 2024; ↑UNESCO, 2024a).

AI integration in education is also gaining momentum in Pakistan, particularly in lesson planning and instructional support. Programmes like Khan Academy[7] and initiatives like LUMSx's 'AI in the Classroom'[8] course train K-12 educators in AI-powered lesson planning. However, infrastructure gaps, teacher training, and AI localisation challenges persist. AI tools for teachers remain concentrated mainly in well-resourced institutions, with little evidence available on local adoption challenges or potential solutions. The effectiveness of AI-generated lesson plans varies between urban and rural schools, influenced by internet access, digital literacy, and teacher training. Research highlights that rural schools in Pakistan, India, and the US struggle with AI adoption, requiring targeted policy interventions (↑Kaufman et al., 2025; ↑Khurshid et al., 2024). The RAND Corporation further warns that schools with a higher proportion of students from low-income backgrounds are less likely to receive AI training and guidance, exacerbating existing disparities in AI adoption (↑Kaufman et al., 2025).

---

[6] See https://shikshacopilot.in/. Retrieved 27 March 2025.

[7] See https://khanacademypakistan.org/. Retrieved 27 March 2025.

[8] See https://lumsx.lums.edu.pk/ai-in-the-classroom/. Retrieved 27 March 2025.

Future research should focus on ethical AI deployment, policy interventions for teacher training, and localised equitable adaptations, ensuring protection of learners and teachers, and a human-centered approach to AI-generated content (↑UNESCO, 2024b). These areas directly align with our research questions, which examine the quality, effectiveness, and adaptability of AI-generated lesson plans, their impact on pedagogical practices, and teacher perceptions of their usability. Centring teacher agency and policy-driven frameworks will position AI as a supportive tool, ensuring its sustainable and equitable integration in education while addressing challenges in both urban and rural settings.

3. Methodology

We adopted a mixed-methods research design for our rapid evaluation, combining quantitative metrics with qualitative insights for a comprehensive analysis. Triangulation enhances reliability by validating findings across diverse data sources. This flexible framework allows for methodological adaptations based on emerging insights, ensuring alignment with project dynamics and stakeholder needs.

3.1. Overview of evaluation activities

The study assessed the usability, effectiveness, and pedagogical impact of AI-generated lesson plans through a combination of quantitative and qualitative methods, which included:

- A Teacher survey capturing insights on lesson plan effectiveness, teacher perceptions, and shifts in teaching practices.

- Key informant interviews gathering insights from subject-matter experts and a school administrator on implementation challenges and curriculum alignment.

- Teacher focus group discussions exploring teacher experiences with the centrally provided lesson plans, quality and usability concerns, and requisite changes in practice.

- An expert review[9] of the lesson plans evaluating the AI-generated lesson plans by developing a customised evaluation framework synthesised by combining elements from established educational frameworks to assess quality, adaptability, and instructional effectiveness.

The mode of communication for these data-collection tools was bilingual, with both English and Urdu used to enhance communication and ensure clarity in inquiry and responses.

3.2. Teacher survey

A comprehensive survey was administered to a total of 160 teachers actively using NIETE's lesson plans across various grade levels and subjects. The teachers were from both rural and urban schools and had a range of experience; some had over ten years of experience, while others were just

---

[9] The review was conducted by a senior academic expert in global education, and a former technical lead for science at the National Curriculum Council.

beginning their teaching careers. Teachers participated voluntarily, and were engaged through the joint outreach of area education officers and head teachers, instead of the survey being mandated by their employers, which may have affected the quality and integrity of responses.

The survey aimed to capture both quantitative and qualitative dimensions of user experiences, offering a nuanced understanding of the intervention's impact. Quantitative responses were analysed through statistical methods, while insights generated through open-ended responses were thematically coded to extract key insights. Questions were designed to assess teachers' perceptions of the centralised repository of lesson plans, focusing on their effectiveness, usability, and adaptability, aided by a structured comparison with their own lesson plans and the process of developing them. Building on the thematic framework employed in other evaluation methods, including focus group discussions, the survey followed the same structured approach, addressing key themes such as implementation challenges, contextual relevance, and pedagogical effectiveness.

The survey explored the impact of these AI-generated lesson plans on:

- Curriculum alignment

- Student engagement

- Differentiated instruction

- Teaching methodologies and pedagogical approaches

- Assessment strategies.

Additionally, the survey examined teachers' confidence levels in using the lesson plans, their comparisons with their self-created lesson plans, and identified barriers to effective implementation, such as infrastructure constraints and accessibility of support materials. It also sought feedback on the structure of the lesson plans, ease of transitioning between lesson components, and insights into their implementation within classrooms.

Apart from questions that gauged the intrinsic quality of lesson plans according to the teacher's perceptions and experiences, the survey delved further to investigate external factors influencing the adoption and effectiveness of NIETE's lesson plans. A key objective was to identify gaps hindering effective integration for teachers, ensuring that findings not only highlighted strengths but also provided insights into potential actionable improvements to the programme.

The survey was piloted with six teachers before dissemination to gauge the tool's effectiveness.

3.3. Key informant interviews

Key informant interviews provided in-depth qualitative insights from education experts, administrators, and subject-matter specialists, offering a high-level perspective on the development, evolution, and integration of AI-generated lesson plans. These interviews assessed pedagogical value, alignment with curricular standards, implementation challenges, and scalability within diverse educational contexts.

3.3.1. NIETE subject-matter experts and lesson plan developers: Key informant interviews 1 and 2

Interviews with subject-matter experts focused on evaluating the development and quality assurance processes of the AI-generated lesson plans. Participants included language, maths, and science specialists, along with a curriculum design lead with expertise in instructional design and educational technology. The semi-structured interviews explored several aspects related to the quality of lesson plans, including curriculum alignment, instructional coherence, and the effectiveness of AI-generated content. These interviews provided a critical framework for evaluation, allowing developers' perspectives to be juxtaposed with expert analysis of the lesson plans.

3.3.2. School administrator: Key informant interview 3

An interview with a Federal Directorate of Education school administrator examined pre-existing conditions, available resources, and instructional practices, establishing a comparative framework for evaluating the integration of centralised AI-generated lesson plans. This discussion focused on resource alignment, pedagogical approaches, and institutional challenges influencing adoption. By capturing data on existing practices and attitudes, the interview informed recommendations for optimising teacher training and refining best practices for implementation.

3.3.3. Independent expert: Key informant interview 4

An independent expert reviewer provided comparative insights on the AI-generated lesson plans against global and local educational standards. Conducted remotely, this interview assessed the objective quality of the NIETE lesson plans, examining how they compared to established federal, provincial, and international benchmarks. This expert perspective is critical

in evaluating scalability and ensuring the lesson plans meet rigorous pedagogical standards before broader implementation.

By synthesising insights from subject-matter experts, administrators, and independent reviewers, these key informant interviews strengthened the evaluation framework, ensuring a rigorous, multidimensional assessment of NIETE's AI-generated lesson plans. The findings can inform targeted refinements, enhancing pedagogical effectiveness, curriculum alignment, and the feasibility of scaling the lesson plans across diverse educational settings.

3.4. Teacher focus group discussions

Focus group discussions provided peer-driven insights into the usability, instructional effectiveness, and contextual relevance of NIETE's AI-generated lesson plans. The focus group discussions explored teacher experiences with these lesson plans compared to their previously used self-planning methods, identifying key advantages, challenges, and areas for refinement.

3.4.1. Objective

The focus group discussions aimed to gather nuanced insights from teachers regarding their experiences with the new central repository of lesson plans, particularly in terms of lesson quality, relevance, deliverability, adaptability, and instructional effectiveness. They also sought to compare the current delivery process, quality, and effectiveness of these AI-generated plans with teachers' previous self-planning methods. Additionally, the focus group discussions explored the extent to which teachers followed the AI-generated lesson plans as scripted, identified the challenges they faced in implementing them, and highlighted both strengths and areas for improvement.
Two structured focus group discussions were conducted to gather insights from teachers on their experiences with AI-generated lesson plans. The first focus group discussion included 12 early primary teachers (Grades 1–3), while the second was intentionally limited to six upper primary teachers (Grades 4–5) to enhance discussion effectiveness and ensure deeper engagement. This approach facilitated more structured contributions and minimised conversational overlap. Each session lasted between 90 and 120 minutes.

3.4.2. Thematic framework

The discussions followed a structured guide, covering key areas such as instructional quality, assessment strategies, implementation fidelity, and

the comparative effectiveness of scripted versus self-planned lessons. These themes were designed to elicit teacher-driven insights, allowing participants to share real-world experiences, challenges, and classroom adaptations. A trained moderator facilitated both sessions, ensuring balanced participation, and guided the conversation to capture diverse perspectives. This thematic framework, which was used in the development of the structured guide, was employed consistently across the evaluation activities, leading to a seamless triangulation process.

3.4.3. Data collection

Data collection involved detailed field notes, taken with participant consent, to document key discussion points, emerging themes, and non-verbal cues relevant to classroom implementation. A thematic analysis was conducted using a structured framework that categorised insights into five key areas:

1. Instructional quality and adaptability

2. Assessment strategies

3. Student engagement and activities

4. Curriculum alignment and contextual relevance

5. Teacher workload and overall feasibility.

Responses were further analysed to capture contextual variations, based on the teachers' years of experience and whether they taught in rural or urban schools. A comparative lens was applied to assess fidelity in implementing scripted lesson plans, identifying areas where teachers deviated from the structure and exploring the reasons behind these adjustments.

3.5. Expert review of lesson plans

A structured and objective review was conducted to evaluate the AI-generated lesson plans, ensuring an objective assessment of content quality, instructional design, and alignment with pedagogical best practices. Education experts evaluated Taleemabad's AI-generated lesson plans using a customised multi-criteria evaluation framework grounded in established educational models, focusing on curriculum alignment, instructional effectiveness, accessibility, and teacher usability. The evaluation framework incorporated Bloom's Taxonomy to assess cognitive rigour. Universal Design for Learning principles were applied to evaluate

adaptability and accessibility, ensuring that lesson plans accommodated diverse learning needs.

3.5.1. 'Zoom in, zoom out': Our evaluation approach method

We used a 'zoom in, zoom out' approach to comprehensively explore the topics or chapters of the textbooks from which the sampled lesson plans were selected. This method allows for a broader contextualised analysis as well as a detailed in-depth evaluation of the lesson plans.

Figure 3.1. '*Zoom in, zoom out' approach to evaluation*

For this purpose, two sets of reviews were carried out by subject-matter experts. The first review, for language subjects and mathematics, was conducted by a senior academic expert overseeing academic programmes at two of Pakistan's largest government schools. With a PhD from a leading international university, their research focuses on global education discourses in Pakistani social studies textbooks. The review for general knowledge (GK) and science was conducted by a former technical lead for science at the National Curriculum Council, with expertise in instructional design, science, technology, engineering, and mathematics (STEM) education, and educational technology. This approach provided diverse perspectives, combining expertise in curriculum development, instructional design, and education policy to assess the quality of lesson plans.

3.5.2. Evaluation framework

The expert review followed a structured evaluation framework to assess the effectiveness of AI-generated lesson plans across multiple dimensions. The assessment examined curriculum alignment, ensuring lesson

objectives adhered to national standards and subject-specific pedagogical models such as Concrete-Pictorial-Abstract for maths, and 5E Instructional Model and Inquiry-Based Learning for science (⬆Bybee et al., 2006). The Backward Design Framework was applied to evaluate how well lesson plans aligned with curricular goals, ensuring that learning outcomes guided instructional strategies and assessments (⬆Wiggins & McTighe, 2005). Instructional design was evaluated based on adherence to the Gradual Release of Responsibility model (⬆Pearson & Gallagher, 1983), with a focus on how well lessons transitioned from explanation to guided and independent practice (⬆Fisher & Frey, 2008). Content quality was assessed for clarity, conceptual depth, logical sequencing, and factual accuracy. At the same time, adaptability and differentiation were reviewed to determine how effectively lessons accommodated diverse learning needs, including scaffolding for struggling learners and extension activities for advanced students. The framework also measured student engagement, analysing the inclusion of real-world examples, inquiry-driven tasks, and opportunities for discussion and collaboration. Teacher usability was a key consideration, with experts assessing whether lesson structures were intuitive, well sequenced, and easy to implement without adding to teacher workload. Assessment integration was examined to ensure the presence of formative and summative evaluation strategies aligned with Bloom's Taxonomy and other assessment models. Finally, accessibility and inclusivity were evaluated to confirm that lesson plans used inclusive language, avoided cultural or gender biases, and supported learners with diverse needs.

3.5.3. Lesson plan selection process

Lesson plans were selected using stratified random sampling. This method was chosen to ensure fair representation across subjects and grade levels. This approach aims to achieve a balanced evaluation of different instructional strategies and their effectiveness.

Lesson plans were uploaded and categorised on the NIETE app by subject and grade level. A proportional number of lesson plans were randomly selected from each category to ensure diverse representation. Revision lesson plans were not selected as part of this review.

Table 3.1. presents lesson plans selected for evaluation using the method mentioned above. The topic numbers refer to the textbook chapters from which these lesson plans were picked.

Table 3.1. *Sample lesson plans for expert review*

	Mathematics	**GK/Science**	**Urdu**	**English**
Grade 1	Topic 4—Lesson 3	Topic 8—Lesson 2	Topic 1—Lesson 26	Topic 2—Lesson 2
Grade 2	Topic 8—Lesson 1	Topic 10—Lesson 1	Topic 6—Lesson 2	Topic 6—Lesson 1
Grade 3	Topic 6—Lesson 4	Topic 8—Lesson 2	Topic 10—Lesson 5	Topic 11 – Lesson 5
Grade 4	Topic 5—Lesson 5	Topic 5—Lesson 2	Topic 14—Lesson 2	Topic 8—Lesson 3
Grade 5	Topic 3—Lesson 2	Topic 5—Lesson 5	Topic 19—Lesson 5	Topic 4—Lesson 5

A structured rubric-based approach was used, with each criterion rated on a five-point scale, ensuring consistent, evidence-based evaluation and quantifiable insights into lesson quality. The final review provided targeted recommendations to refine lesson plan design and ensure scalability and pedagogical effectiveness. By integrating insights, the expert review provided a clear assessment of the strengths and areas for improvement in AI-driven lesson planning.

3.6. Participants and sampling strategy

The sampling strategy employed in this evaluation was designed to gather data from diverse educational contexts where Taleemabad's AI-generated lesson plans are implemented through the NIETE programme. A stratified purposive sampling method was used to include schools across both urban and rural sectors and settings. This approach ensured the inclusion of varied socioeconomic and geographic characteristics within the Islamabad Capital Territory region, which are critical for assessing the lesson plans' effectiveness and adaptability.

3.6.1. Strata definition

The sample was divided into two strata based on school location and resource availability:

- Urban schools: Representing two relatively well-resourced sectors with greater access to technology and infrastructure.

- Rural schools: Representing four underserved regions with limited access to educational resources and digital infrastructure.

The sample size was determined following best practices for focus group discussions, balancing the depth and breadth of inquiry with budgetary and logistical constraints. Each focus group discussion included a minimum of three to six teachers per stratum to ensure meaningful discussion and comparative insights.

Additionally, to enhance statistical validity, a quantitative survey was disseminated to all teachers using NIETE's AI-generated lesson plans.

3.6.2. Teacher selection

Within each stratum, purposive sampling was used to select teachers based on specific criteria. Participants included those who were either 'mostly on schedule' (had completed more than 60% of the allotted lessons in a week), 'moderate users' (had completed 20–60% of the allotted lessons in a week), or 'minimally engaged' with the lesson plans (0–20% completion), with a lower cut-off threshold requiring teachers to have opened and read at least three lesson plans on the NIETE platform.

The sample also ensured diversity across subjects and grade levels, recognising that most primary-level teachers teach multiple subjects rather than being restricted to either language or technical disciplines. Furthermore, teachers with varying levels of experience and familiarity with educational technology were included to capture a broad spectrum of perspectives on implementation and usability.

3.6.3. Focus group discussions

For the focus group discussions, anonymised teacher data was used to ensure confidentiality. The first focus group discussion comprised teachers from Grades 1 to 3, facilitating peer-driven insights into early primary teaching experiences with subject-specific AI-generated lesson plans. The second focus group discussion included teachers from Grades 4 to 5, capturing insights from upper primary educators.

Table 3.2. *Sample selection of teachers for the focus group discussions (FGDs)*

FGD	No. of teachers	Grades	Rural	Urban
1	12	1–3	6	6
2	6	4–5	3	3

3.7. Data analysis

To evaluate the effectiveness and adaptability of Taleemabad's AI-generated lesson plans, a systematic data analysis approach was employed, incorporating both qualitative thematic analysis and triangulation methods to ensure validity, reliability, and depth of insight. By identifying key trends and patterns, detecting outliers and discrepancies, and extracting actionable insights, the analysis provides a comprehensive understanding of the quality of these lesson plans and their impact on teacher load and pedagogical practices.

3.7.1. Qualitative analysis and thematic coding

Qualitative data from key informant interviews and focus group discussions underwent thematic analysis using a structured coding framework. The process began with a thorough review and synthesis of field notes to identify recurring themes and patterns. Data was then categorised into broader themes, such as adaptability, engagement, assessment effectiveness, and contextual barriers. A triangulation matrix was developed to cross-validate findings from focus group discussions, survey, key informant interviews, and expert evaluations, ensuring alignment with the core evaluation themes. This mapping process strengthens the credibility of insights by revealing patterns and inconsistencies across the qualitative and quantitative data.

3.7.2. Integration of platform data

Platform data was also integrated into the analysis, providing teacher-specific usage insights that contextualised survey and interview findings. The platform data included teacher names, schools, years of experience, subjects taught, frequency of lesson plan usage, and sector classification (urban or rural). This data allows for a more granular understanding of how teachers with different backgrounds and teaching environments engaged with the NIETE lesson plans, offering insights that provide critical context for understanding data, especially from the survey and focus group discussions.

3.7.3. Triangulation

Multiple triangulation methods were applied to ensure the reliability and depth of findings.

Method triangulation was used to cross-reference qualitative data from focus group discussions and key informant interviews with quantitative survey responses and statistical analysis.

Data triangulation incorporated various sources, including interviews, surveys, document reviews, and platform data, to capture a comprehensive picture of teachers' experiences.

Investigator triangulation engaged multiple analysts to minimise bias and ensure a balanced interpretation of findings.

3.7.4. Thematic evaluation framework

A thematic evaluation framework guided the synthesis of insights across all evaluation activities. This framework structured the analysis around eight key dimensions.

Each dimension was assessed using specific evaluation criteria and indicators, ensuring a rigorous and systematic approach to understanding the impact of the AI-generated lesson plans. The framework also served as the basis for expert reviews, which provided an additional layer of validation and quality assessment.

Table 3.3 below outlines the evaluation criteria and corresponding indicators used to assess the quality and effectiveness of the AI-generated lesson plans.

Table 3.3. *Evaluation criteria and their indicators*

	Evaluation criteria	Indicators
1	Alignment with curriculum and standards	■ Lesson objectives align with the national curriculum. ■ Content is mapped accurately to prescribed textbooks and learning standards. ■ Lessons adhere to subject-specific pedagogical approaches (e.g., Concrete Pictorial Abstract, 5E Instructional Model, Inquiry-Based Learning).
2	Instructional design and pedagogical coherence	■ Follows the Gradual Release of Responsibility model effectively. ■ Science lessons integrate 5E Instructional Model & Inquiry-Based Learning elements appropriately. ■ Maths lessons incorporate Concrete-Pictorial-Abstract progression effectively.

	Evaluation criteria	Indicators
3	Clarity, depth, and accuracy of content	■ Explanations are clear and logically structured. ■ Key concepts are well articulated with appropriate depth. ■ Examples and illustrations enhance understanding. ■ Content is factually accurate and free from errors.
4	Adaptability and differentiation	■ Provides strategies for different learning paces and student needs. ■ Includes scaffolding for struggling learners. ■ Allows for extension activities for advanced learners. ■ Accommodates diverse classroom contexts (e.g., rural vs. urban settings).
5	Engagement and student-centred learning	■ Lesson activities encourage active participation. ■ Use of real-world examples and contextualised scenarios. ■ Opportunities for student discussion and collaboration. ■ Inquiry-driven tasks in science and problem-solving in maths.
6	Ease of use for teachers	■ Lesson structure is intuitive and easy to follow. ■ Clear instructions and guidance for implementation. ■ Logical sequencing of activities. ■ Does not create an additional burden on teachers.
7	Assessment integration	■ Includes well-structured formative assessment opportunities. ■ Questions align with Bloom's Taxonomy and other assessment frameworks. ■ Provides feedback mechanisms to guide student progress. ■ Supports summative assessment for teachers.
8	Accessibility and Inclusivity	■ Uses inclusive language and examples. ■ Avoids cultural or gender biases.

Evaluation criteria	Indicators
	■ Supports students with diverse learning needs.

By integrating thematic analyses, platform data insights, and a robust triangulation approach, this evaluation provides a high-confidence assessment of Taleemabad's AI-generated lesson plans, highlighting both strengths and areas for improvement. The findings offer concrete recommendations for optimising lesson plans, improving teacher support mechanisms, and addressing barriers to adoption.

3.8. Ethical considerations

3.8.1. Informed consent

All participants were provided detailed information on the purpose of the assessment, research methods, and confidentiality protocols. In addition, they were provided with the opportunity to ask any questions or concerns. Verbal or written consent was obtained to ensure ethical compliance. Participants were explicitly informed that their contributions would be used solely for research and programme improvement. Focus group discussion participants also signed consent forms before the start of the discussion.

As noted, the study's design necessitated the research team not informing teachers that the NIETE lesson plans were AI-generated. This constraint assured unbiased feedback and authentic engagement with the materials.

3.8.2. Confidentiality

All data was anonymised, and in order to protect participant privacy, no personally identifiable information was recorded. Pseudonyms were used in field notes and reports to prevent direct identification. Any references to participants' organisations or locations were included only as far as they related to the study and without specific city, district, or school names. The survey carried a clear notice ensuring anonymity, and facilitators of focus group discussions also ensured participants knew complete anonymity and confidentiality would be maintained. The study adhered to EdTech Hub's data protection policies, ensuring strict confidentiality and ethical compliance.

3.8.3. Voluntary participation

Participation was entirely voluntary, with individuals free to withdraw at any time without consequence. This was a key feature of the research, ensuring high-quality data was collected from teachers who were self-motivated to provide feedback, rather than responses obtained due to a directive from the school administrators. We contacted teachers directly for the survey and focus group discussions, rather than through their school authorities. Knowing this was an independent inquiry allowed participants to give unbiased feedback on not only the lesson plans, but also extraneous factors related to school infrastructure or issues related to administration.

3.9. Limitations

While this rapid evaluation provides valuable insights into the effectiveness of AI-generated lesson plans, we acknowledge certain methodological limitations.

3.9.1. Study design and methodological constraints

The first limitation pertains to the study's design. Since no baseline data was available for comparison, the study did not employ an experimental or quasi-experimental design. As a result, while we incorporate both qualitative insights from teachers and quantitative data from the survey and expert rubric evaluations, we can only establish causal relationships between the AI-generated lesson plans and any observed changes in teaching practices or student engagement with limited confidence.

3.9.2. Scope of participant access

A key limitation of this study is the restricted scope of participant access, as data collection was limited to teachers, school administrators, and developers of the AI-generated lesson plans. While these stakeholders provide valuable insights into the implementation and usability of the lesson plans, the study does not include an evaluation of impact on student learning, which could have offered a more holistic understanding of the broader impact of these lesson plans on learning experiences and educational outcomes. Additionally, while developers provided information on how the lesson plans were created, the study did not assess the AI generation process itself, limiting our ability to analyse the computational efficiency, decision-making processes, and technical performance of the AI model.

3.9.3. Teacher awareness of AI-generated content

Another significant limitation stems from the fact that teachers were not informed that the lesson plans they were using were AI-generated. This restricted the study's ability to explore teachers' perceptions of AI in education, including any biases, concerns, or expectations they may have had regarding AI-generated content. Without this knowledge, teachers could not reflect on whether AI-generated lesson plans differed from human-created ones in ways they might consciously recognise or value.

However, this lack of awareness also provided a unique methodological advantage: it allowed for an unbiased comparison between the AI-generated lesson plans and the previously used guidebook-based lesson planning process. Because preconceived notions about AI did not influence teachers, their feedback was based purely on their experience of using the lesson plans in their classrooms.

This study therefore offers a teacher-centred evaluation of a centralised repository of AI-generated lesson plans, assessing their practical impact on instructional delivery. Furthermore, it extends this evaluation to consider the potential of AI-powered educational materials for scalability and accessibility in low-resource settings and LMICs, where high-quality, adaptable lesson planning tools are critically needed.

3.9.4. Implementation-related challenges

The research team also encountered some implementation-related limitations during the study. In the focus group discussions, an audio recording equipment malfunction resulted in the loss of recorded data. However, detailed note-taking and direct observation ensured that key insights were still captured.

Another challenge was the number of participants in the focus group discussions. The first session, which included 12 participants, provided a broad range of perspectives but also made it challenging at times to ensure equal participation with some overlaps and disruptions. To address this challenge, the second focus group discussion was conducted with just six participants, which significantly improved clarity in responses and allowed each teacher sufficient time to share detailed insights.

To ensure that the perspectives of the remaining six participants were still incorporated, they were invited to participate in the survey pilot, which generated valuable insights. This process led to the inclusion of key questions related to ease of use, such as internet connectivity challenges,

based on teachers' experiences with the practical aspects of implementation.

3.9.5. Survey participation and sample size

Lastly, rather than implementing a design where teachers were mandated by their employers to complete the survey, which could have increased response rates but potentially influenced the nature of their responses, the study opted for a voluntary participation model. A key limitation this posed was the relatively modest sample size of 160 respondents. However, this approach ensured teachers engaged willingly and provided more candid feedback.

While some of these limitations narrow the study's scope, they also ensure that the evaluation focuses on practical usability and instructional quality rather than being influenced by attitudes toward AI. Future research could build on these findings by incorporating broader stakeholder perspectives and explicitly examining teachers' perceptions of AI-generated educational materials.

4. Findings

This study's findings are derived from a comprehensive analysis of qualitative and quantitative data collected through multiple sources. Using matrix triangulation, we identified key themes that emerged from focus group discussions, a teacher survey, key informant interviews, and expert reviews of the lesson plans. The focus group discussions, survey, and key informant interviews primarily focused on comparing the new NIETE lesson plans with the previous lesson planning process—assessing their ease of use, perceived effectiveness, and overall quality. Meanwhile, the expert review provided an objective evaluation of the lesson plans based on a structured framework developed for this study.

This section synthesises the collected data into emergent themes, presenting a holistic understanding of how the AI-generated lesson plans are perceived and assessed across different stakeholder groups. By integrating insights from teachers' experiences and expert evaluations, the findings aim to provide a nuanced perspective on the strengths and areas for improvement in the NIETE lesson plans.

While these themes focus on the comparative quality and implementability of Taleemabad's AI-powered lesson plans, meta-themes such as contextualisation, teacher confidence, and adaptability appear across all of these and are discussed in the conclusions and recommendations in Section 5 to inform actionable insights and improvements.

4.1. Adoption

In evaluating the ease of use and adoption of these lesson plans, an interesting contradiction emerged between self-reported usage and the platform data on usage. To gauge actual adoption, we looked at the teacher data from the NIETE platform and compared it against the data on usage and ease of implementation gathered from the teachers through the survey and focus group discussions.

Using a cutoff threshold of teachers who had accessed at least three NIETE lesson plans on the platform, the analysis of the usage data from the NIETE platform indicates the breakdown presented in Figure 4.1 for the 3,423 teachers who had opened at least three NIETE lesson plans.

Figure 4.1. *Breakdown of 'on schedule' and 'not engaged' teachers*
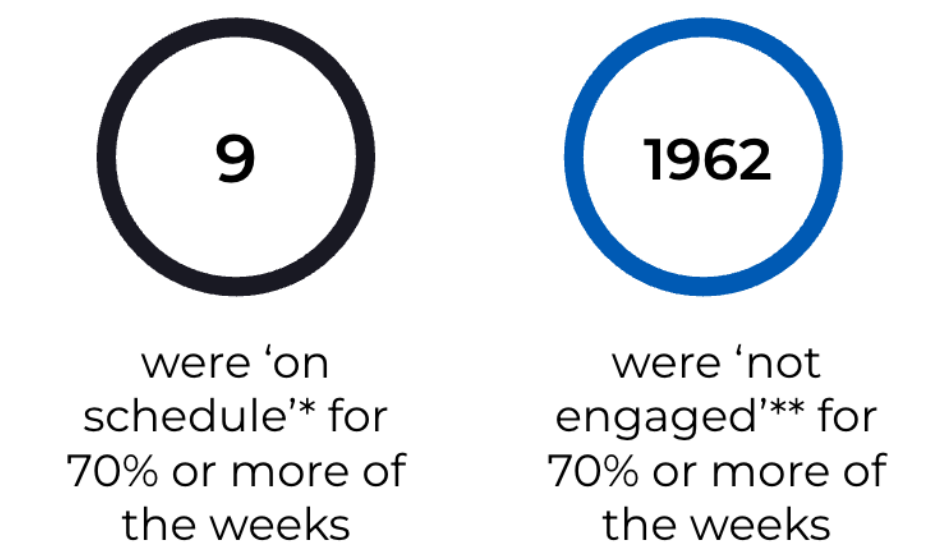
Notes: *'on schedule' means teachers used above 60% of the allotted lesson plans in the given time.
**'not engaged' means teachers did not use the allotted lesson plans.

Data from the survey, as presented in Figure 4.2 shows that 62.5% of the teachers used the lesson plans daily while 20% more used it at least once a week.

Figure 4.2. *Survey question: How frequently do you use the new NIETE lesson plans?*
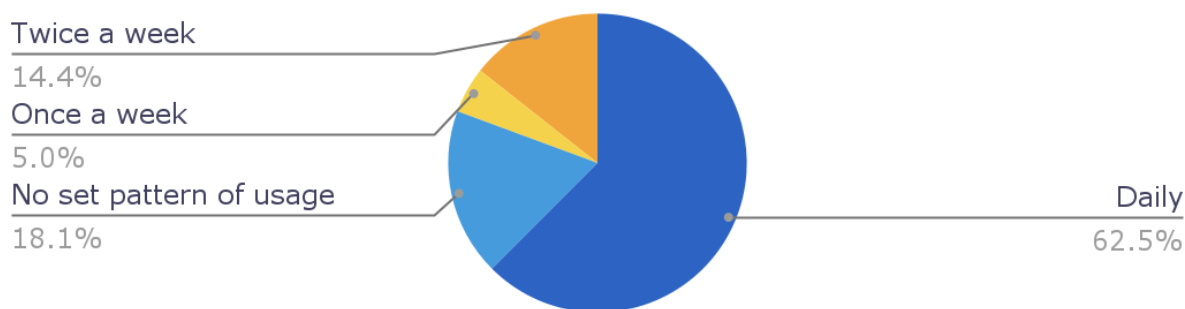
This contradiction provides insights on teachers' perceptions, confidence, and attitudes towards the new lesson plans.

4.2. Ease of use and implementation

In order to further contextualise and understand the data on adoption, it was paired with the data gathered to evaluate the ease of use and implementation of the lesson plans.

For our analysis, we chose the following indicators to assess the ease of use and implementation of these lesson plans across all the data sources.

Indicators for ease of use and implementation include:

- Clear instructions and guidance for implementation.

- Does not create an additional burden for teachers.

- Accommodates diverse classroom contexts (e.g., rural vs. urban settings, varying classroom sizes).

4.2.1. Teachers' perspectives (survey and focus group discussions)

To gather the teachers' perspectives on the ease of implementing the lesson plans in their classrooms, we structured the questions to focus on:

- Ease in preparation

- Ease in delivery with the given resources

- Ease in delivery within the given class period.

The survey responses all came out in favour of the NIETE lesson plans compared to their previous, self-created lesson plans and their development process (see Figure 4.3 and Figure 4.4).

Figure 4.3. *Survey responses: Ease in preparing lesson plans (LPs)*
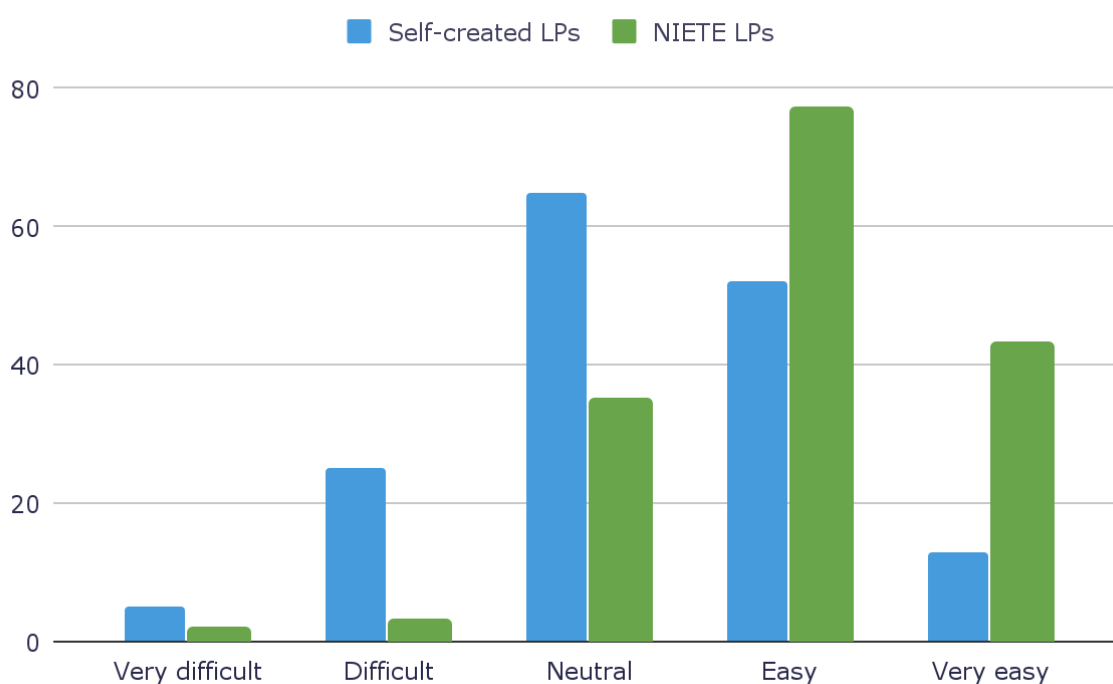
Figure 4.4. *Survey responses: Ease in delivering lesson plans with given resources*
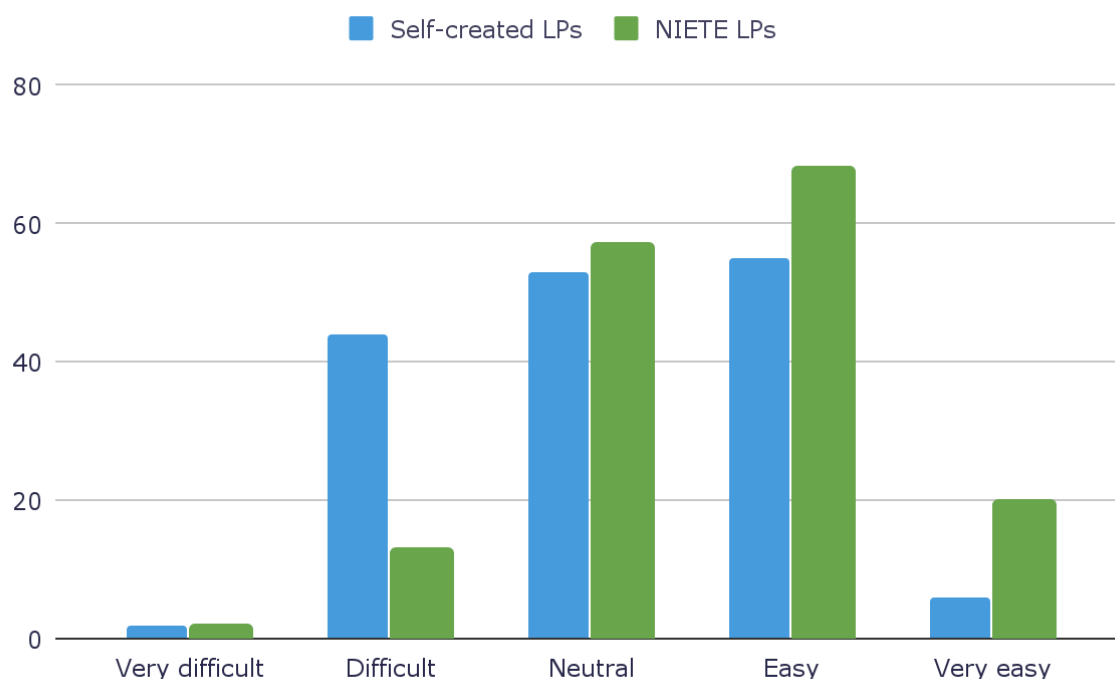
The results presented in Figures 4.3 and 4.4 above indicate that teachers perceived the NIETE lesson plans to be easier to implement than their own self-developed lesson plans.

However, detailed discussions and insights from the focus group discussions and the qualitative responses on the survey suggest that the teachers were struggling with the rigid time-bound structure of the NIETE lesson plans during implementation. Figure 4.5 presents key relevant insights.

Figure 4.5. *Synthesised highlights from survey responses: Time management with the given lesson plan*

The lesson plans are too rigid to account for real classroom sizes and structures, and the challenges that they present in terms of time-management.	Teachers highlighted the importance of ensuring activities fit within the classroom period considering implementation in large class sizes.	The tightly scripted and structured nature of the lesson plans make it harder to improvise and/or modify according to the needs of the class or specific situations.

The quantitative data reflects teachers' appreciation of the NIETE lesson plans in terms of ease of use as compared to their previous methods and lesson plans, and the qualitative data sheds light on where improvements can be made to further facilitate the teachers in delivering quality lessons using AI-generated lesson plans.

4.3. Quality of the lesson plans

The study's main purpose was to evaluate the quality of the AI-generated lesson plans and compare them to the existing or previous lesson plans created by teachers. The study reveals that there were no defined processes, standards, or models for the development and use of lesson plans before the AI-generated and centrally developed lesson plans were introduced through the NIETE programme. Insights from the key informant interviews and focus group discussions revealed that the quality, consistency, as well as the existence of lesson plans depended on varying factors such as oversight of the head teacher or principal, experience and expertise of the teacher, personal approach of the teachers to lesson planning, and availability of time and resources. This shifted the focus of the research from directly comparing lesson plans to examining the differences between the development and implementation processes of the previously used lesson plans and the ease of preparation and quality of the AI-generated lesson plans.

In order to evaluate the quality of the lesson plans, a specialised framework was developed to assess different elements of the lesson plans based on their composite design models, catering to the variations in the design of the lesson plans based on subject areas. The findings are presented under the following themes:

- Curricular alignment

- Instructional design and pedagogical coherence

- Assessment strategy and integration

- Differentiation and adaptability to student needs

- Engagement and classroom activities.

4.3.1. Curricular alignment

One major question posed by the AI-generated lesson plans — or, in fact, any set of lesson plans — is their relevance to the designated curriculum. For our study, we assessed the curricular alignment of these lesson plans to the learning objectives and standards set in the National Curriculum of

Pakistan (2023)[10] through analysis of the generation and quality assurance process, independent expert review, and teacher perception.

Indicators of curricular alignment include:

- Lesson objectives clearly align with the national curriculum

- Content is mapped accurately to prescribed textbooks and learning standards

- Lessons adhere to subject-specific standards

Curricular alignment in the development process

The following approaches were used to ensure curricular alignment of the lesson plans during their development.

Previously used lesson plans created by teachers

The teachers followed the textbooks to develop their lesson plans. Data from a key informant interview and the focus group discussions reveals that the teachers leaned on the textbooks to guide curricular alignment and catered to the learning objectives and their depth and scope as reflected in the textbooks.

> "*Teachers are not too familiar with the curriculum. The books are the curriculum for most of them.*"
> (Key Informant Interview 3)

NIETE's AI-powered lesson plans

The development process of the AI-generated lesson plans is fundamentally based on the textbooks in use in the classrooms.

The 'author', who is the subject-matter specialist in charge of developing the lesson plans, takes excerpts from the textbooks and feeds them to the AI model along with prompt restrictions on using only the provided input to develop a lesson plan. The author then reviews the generated lesson plans again to ensure the plans follow the inputs provided to the model, and the author removes or modifies any irrelevant additions. The team lead reviews the final output to ensure misalignments are reduced or removed. As a result, curricular alignment is ensured by only providing the relevant inputs and then reviewing for irrelevant content.

---

[10] See https://pctb.punjab.gov.pk/node/219. Retrieved 8 April 2025.

Teachers' perspectives (survey and focus group discussions)

The survey asked teachers to compare the curricular alignment of the NIETE lesson plans with the curricular alignment of their own lesson plans; Figure 4.6 below highlights a key finding.

Figure 4.6. *Survey responses: Curricular alignment of the lesson plans*

Considering that the teachers develop their own lesson plans using the textbooks provided to them, the fact that 80% teachers consider NIETE's lesson plans to be perfectly or well aligned compared to 74% of the teachers reporting the same satisfaction with their own lesson plans, indicates that they found few issues with the NIETE lesson plans regarding curricular alignment.

Insights from the focus group discussions corroborate these findings, as most teachers agreed that NIETE follows the syllabus/curriculum.

Expert review summary

The expert reviews revealed that most of the sampled lesson plans adhere to the national curriculum guidelines and integrate subject-specific pedagogical approaches. The textbooks guide the lesson plans; even where they have issues or mistakes, the lesson plans follow them.

This reveals that the lesson plans may only be as aligned as the textbooks.

4.3.2. Instructional design and pedagogical coherence

The instructional design quality of the AI-powered lesson plans was one of the key points to consider in assessing the plans. Probing the development process of the plans revealed an ambitious design that incorporates different established and evidence-backed approaches and combines them in an effort to provide effective lesson plans for the different subjects. This approach led us to incorporate pedagogical coherence alongside instructional design in our evaluation framework, to study how the different pedagogical models are combined to develop coherent lesson plans.

Development process

Previously used lesson plans created by teachers

Data from the key informant interviews and focus group discussions reveals that the previous lesson plans, or their development process, did

not incorporate any established pedagogical models. Teachers had the flexibility to use any models that fit their context, but this led to a lack of consistent pedagogical structure in their lesson plans. This resulted in an inconsistent approach across schools that varied with principals/head teachers/teachers' own preferences, experiences, and administrative styles, among other things.

NIETE's AI-powered lesson plans

Taleemabad's team designed the lesson plans based on several established models. They use the overarching design of scripted lesson plans as advised by the World Bank in their report/guidelines for LMICs (↑Akeyampong et al., 2023). The overarching model followed across all grades and subjects is the Gradual Release of Responsibility model with elements taken from the Blooms Taxonomy framework to guide development (↑Krathwohl, 2002; ↑Pearson & Gallagher, 1983).

Further variations are made for the subjects to incorporate relevant tools from corresponding models. For science, elements from the 5E Instructional Model and Inquiry-Based Learning have been incorporated into the lesson plans. For maths, the deployment of the Concrete-Pictorial-Abstract model was chosen to support learning through that established ideology.

Taking into account the given context and approaches used initially as well as for the NIETE lesson plans, the following indicators were developed for curricular alignment and pedagogical coherence as part of the evaluation framework.

- Follows the Gradual Release of Responsibility model effectively.

- Science lessons integrate 5E Instructional Model & Inquiry-Based Learning elements appropriately.

- Maths lessons incorporate Concrete-Pictorial-Abstract progression effectively.

- Explanations are clear and logically structured.

- Key concepts are well-articulated with appropriate depth.

- Examples and illustrations enhance understanding.

- Content is factually accurate and free from errors.

- Lesson structure is intuitive and easy to follow.

Teachers' perspectives (survey and focus group discussions)

Based on the survey responses and focus group discussions, the teachers' perspectives are overwhelmingly positive about the structure provided in these lesson plans. Teachers stated that the clear objectives, structured activities, and formative questions led to greater student engagement. They found that the activities and assessments were integrated effectively, but that they also felt burdened by the volume of tasks in the lesson plans.

Data from the survey shows that about two-thirds of the teachers found the model and structure of the NIETE lesson plans to be extremely or very effective.

Figure 4.7. *Survey question: How effective are the model and structure of NIETE's lesson plans?*

In order to effectively gauge teachers' perspectives on the success of the overarching Gradual Release of Responsibility model, the survey asked about how effective the lesson plans were in encouraging student independence by the end of the lesson (the ultimate goal of the model).

Figure 4.8. *Survey question: How effectively do the NIETE lesson plans encourage student independence by the end of the lesson?*
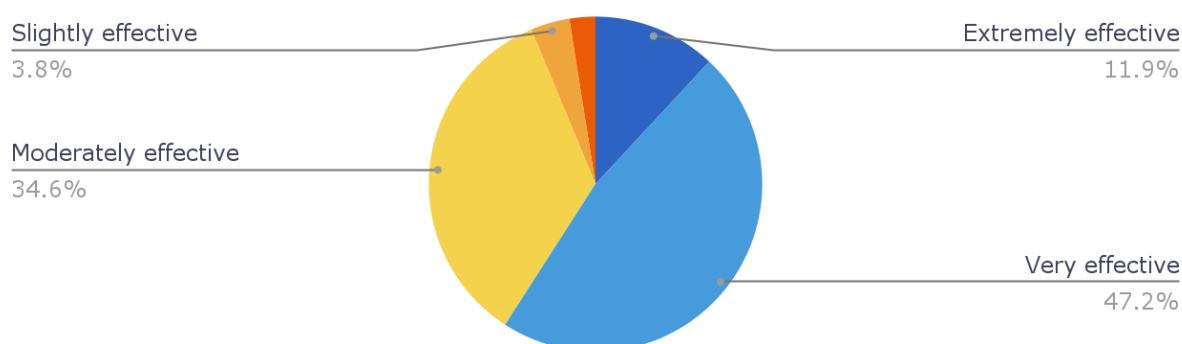
Given the context of teachers not having followed structured pedagogical models, our next point of enquiry was to evaluate how confident they were in following the model incorporated in NIETE's lesson plans. This also provided insights into NIETE's pain point of evaluating implementation fidelity and gauging how capable or confident teachers are in following their lesson plans.

Figure 4.9. *Survey question: How do you find transitioning between the different sections of the NIETE lesson plans?*
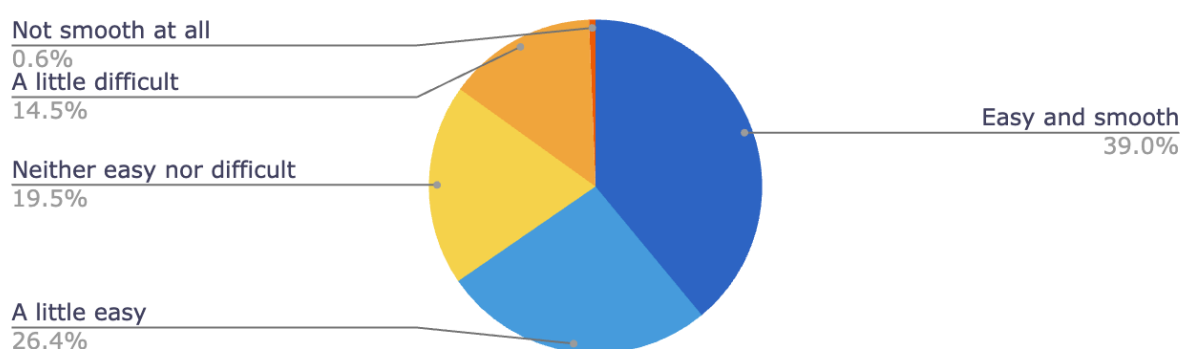
Qualitative feedback in the survey mentions that the structure is clear and easy to follow, but the lesson plans are too rigid for different contexts.

Expert review summary

Most lesson plans followed well-established frameworks, including Gradual Release of Responsibility, which supports student progression from guided to independent learning.

Some lesson plans lacked coherence in instructional flow and moved abruptly between instructional elements such as questions, diagrams, and flashcards without clear connections.

Lesson plans sometimes failed to build foundational understanding before introducing more complex tasks, leading to abrupt transitions that may confuse students.

4.3.3. Assessment strategy and integration

A critical aspect of evaluating the NIETE lesson plans was understanding how they support assessment practices and integrate into existing teaching routines. Lesson plans are fundamentally vehicles of instruction, but evaluation or assessment techniques are also incorporated in lesson plans to ensure student understanding and learning progression. This section explores how the AI-generated lesson plans aligned with assessment strategies, including the ease with which teachers can incorporate formative and summative assessments into their instruction. It also examines how well these lesson plans facilitated student engagement with assessments, and whether they provided sufficient guidance for teachers to evaluate student understanding effectively. Findings in this section draw from teacher feedback, expert evaluations, and key informant insights to assess the extent to which these lesson plans support a coherent and structured approach to classroom assessment.

Evaluation framework indicators

The evaluation framework was used to assess the integration of assessment strategies within the NIETE lesson plans using key indicators that measured their effectiveness in guiding teachers to implement formative and summative assessments. These indicators examine the clarity of assessment instructions, alignment with learning objectives, and the variety of question types included to gauge student understanding. Additionally, the framework was used to evaluate whether the lesson plans provided structured opportunities for feedback and reinforcement, ensuring that assessment is meaningfully embedded in the instructional process.

Indicators for assessment strategy and integration are:

- Includes well-structured formative assessment opportunities.

- Questions align with Bloom's Taxonomy and other assessment frameworks.

- Provides feedback mechanisms to guide student progress.

- Provides teachers with support for summative assessment.

Teachers' perspectives (survey and focus group discussions)

Teachers acknowledged that the assessments in NIETE's lesson plans were more structured and helpful in tracking student progress than their own designs. The frequent usage of formative questions after "every paragraph in the lesson plans" made it easier for teachers to gauge student understanding.

In the survey responses, the teachers rated NIETE's lesson plans' assessment strategy higher than their own lesson plans.

Figure 4.10. *Survey responses: Assessment strategy of the lesson plans*
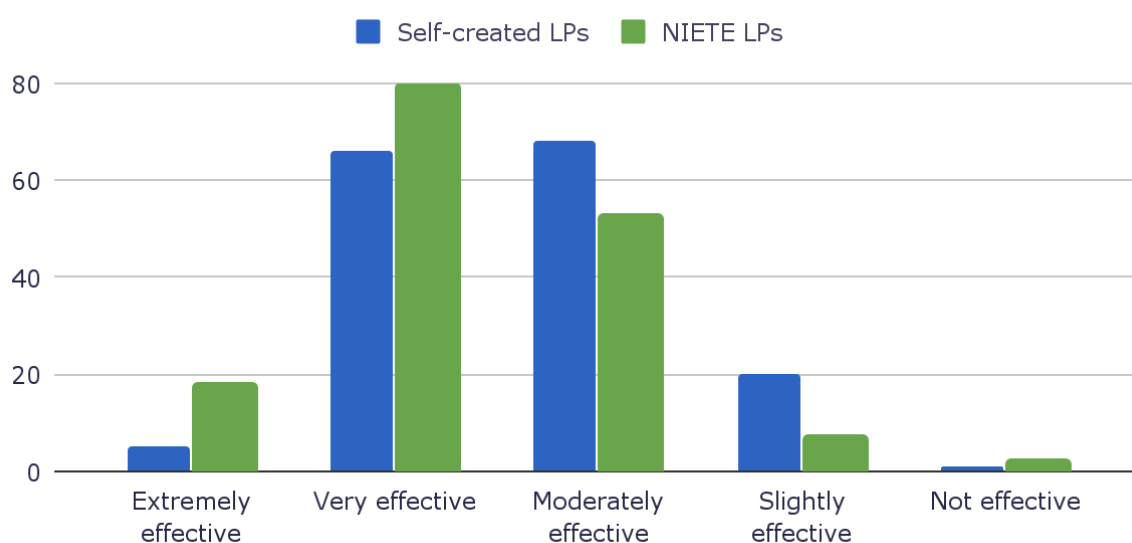
Generally, teachers reported creating assessments as a key challenge when designing their own lesson plans and appreciated the support provided in the NIETE lesson plans. They acknowledge how helpful formative assessments are in understanding student progress, but some teachers struggled with time constraints while carrying out the formative assessments in the provided lesson plans.

Expert review summary

The expert reviews for assessment strategy and integration found a mismatch between instruction and assessment. In most of the sampled lesson plans, assessments test factual recall rather than the skills explicitly taught in the lesson.

The sampled lesson plans also lacked well-structured formative assessments that do not demonstrate sufficient rigour. Rather than incorporating diverse assessment methods, most lesson plans frequently relied on repetitive, simplistic activities such as thumbs-up/thumbs-down responses.

Additionally, many lesson plans did not include higher-order thinking tasks aligned with Bloom's Taxonomy, focusing primarily on rote memorisation rather than promoting analysis, evaluation, or creative problem-solving.

4.3.4. Differentiation and adaptability to student needs

Effective lesson plans must accommodate diverse student needs, ensuring all learners can engage meaningfully with the content. This section

examines the extent to which the NIETE lesson plans incorporate differentiation strategies and adapt to varying student abilities, learning paces, and classroom contexts. Findings in this section draw from teacher feedback, expert evaluations, and key informant insights to assess whether the lesson plans provided sufficient flexibility, scaffolding, and support for a diverse range of learners.

Development process

Previously used lesson plans created by teachers

An analysis of key informant interviews and focus group discussions indicates that the previously used lesson plans and their development process lacked systematic differentiation methodologies. Teachers implemented differentiated pedagogical approaches predominantly based on contextual factors, individual interpretations of learner requirements, and temporal limitations. This practice engendered heterogeneous implementation across the schools, with significant variations contingent upon administrative leadership preferences, professional experiential backgrounds, and individual pedagogical orientations of instructional staff.

Most teachers used group-based strategies to address student differences, and resorted to leveraging peer learning methods to enable stronger students to help the struggling ones.

NIETE's AI-powered lesson plans

NIETE's lesson plans are designed with the different application contexts and the weakest teachers and students in mind. Therefore, the authors tried to incorporate differentiated instructions in the lesson plans and engaged with the team on the field to gather feedback and modify the lesson plans accordingly.

Evaluation framework indicators

The indicators used for differentiation and adaptability for the development of the evaluation framework focused on elements within a lesson plan that facilitate differentiated instruction and learning pace and approaches among the students.

Indicators for differentiation and adaptability were:

- Provides strategies for different learning paces and student needs.

- Includes scaffolding for struggling learners.

■ Allows for extension activities for advanced learners.

Teachers' perspectives (survey and focus group discussions)

The teacher's perspectives in the survey responses and focus group discussions provide mixed views about the differentiation qualities of the NIETE lesson plans.

Results from the survey questions about differentiation in the lesson plans are presented in Figure 4.11 below.

Figure 4.11. *Survey responses: Meeting students' differentiated needs*
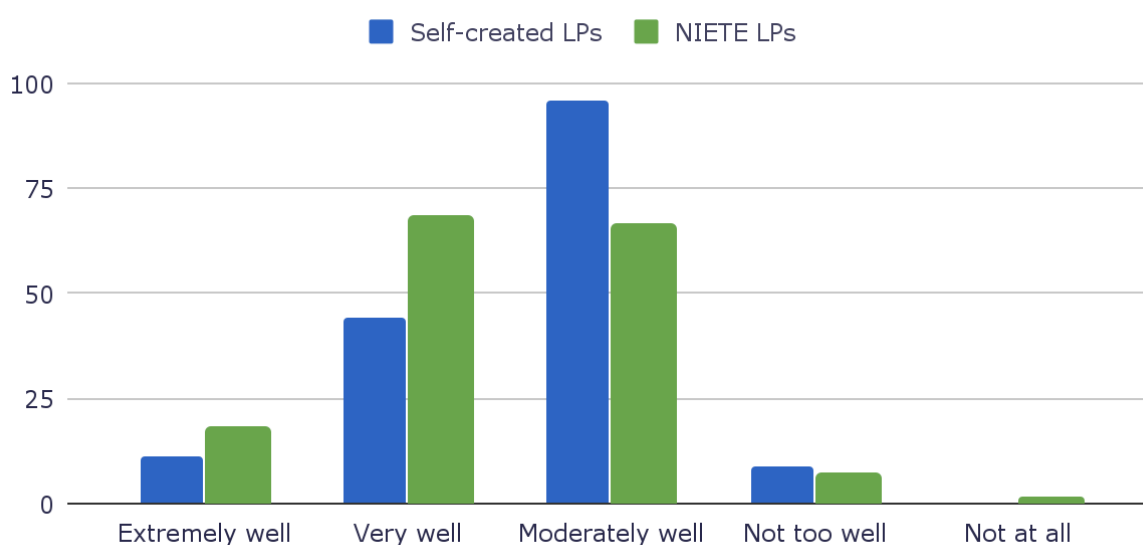
While teachers rated NIETE's differentiation qualities higher than their own lesson plans in the survey, qualitative data from the survey and focus group discussions suggest there is room for significant improvement.

Teachers felt that the NIETE lesson plans lacked flexibility for struggling students. They found that the lesson plans tend not to cater to "the ground realities and students' levels," and that they "should also consider weak students," as reflected in the qualitative responses in the survey.

Detailed focus group discussion data reveals that some teachers needed to modify the NIETE lesson plans to better support struggling students, including additional explanations and time adjustments. They felt that the NIETE lesson plans are effective in differentiation but need to be more adaptive for different student needs in terms of flexibility in the additional time required for instructions for struggling students.

Expert review summary

The expert reviews found inconsistent differentiation strategies in the sampled lesson plans. While some lesson plans provided scaffolding, many lacked clear differentiation strategies for supporting struggling and advanced learners. Furthermore, gaps in conceptual depth and transitions were observed: lesson plans sometimes failed to build foundational understanding before introducing more complex tasks, leading to abrupt transitions that may cause confusion and are insensitive to weak or struggling students.

4.3.5. Engagement and classroom activities

Student engagement is crucial for effective learning, and well-designed lesson plans should incorporate interactive and meaningful classroom activities to foster participation. This section explores how the NIETE lesson plans facilitated student engagement through structured activities, discussions, and hands-on learning experiences. Drawing from teacher feedback, experts' evaluation, and key informant insights, the findings assess whether the lesson plans provided engaging, student-centred activities aligned with pedagogical best practices and enhanced the overall learning experience.

Development process

Previously used lesson plans created by teachers

The data from the key informant interviews and focus group discussions indicates that teachers lacked a centralised collection of classroom activities. As a result, teachers needed to create their own activities, search online resources, or rely on textbook exercises. This additional workload for teachers led to a significantly reduced frequency of classroom activities before the NIETE lesson plans were implemented. In order to develop engaging activities for their lessons, the teachers needed to access online resources or work together to come up with ideas for different topics.

> *"All teachers have a discussion on how to teach about flowers* […] *6 sections we work together (teachers)."*
> (FGD 2 participant)

NIETE's AI-generated lesson plans

Following the pedagogical frameworks and models on which the NIETE lesson plans are based, the NIETE team tried to incorporate contextualised

real-world examples and activities to engage the students. Through the feedback loops incorporated in the NIETE programme as well as their iterative approach towards the development of the lesson plans, their team was able to gather feedback about their proposed activities and modify the lesson plans to incorporate the feedback. As reported in the key informant interviews, the feedback for the activities was about their misalignment with the resources available to some schools. The NIETE team modified their lesson plans to incorporate multiple activity choices for the teachers based on the resources available to them.

Evaluation framework indicators

The indicators identified for engagement and classroom activities as part of the evaluation framework are given below:

- Lesson activities encourage active participation.

- Use of real-world examples and contextualised scenarios.

- Opportunities for student discussion and collaboration.

- Inquiry-driven tasks in science and problem-solving in maths.

Teachers' perspectives (survey and focus group discussions)

The quantitative data from the survey shows that the teachers clearly rate NIETE's lesson plans higher than their own in terms of student engagement (see Figure 4.12).

Figure 4.12. *Survey responses: Student engagement based on the lesson plans*
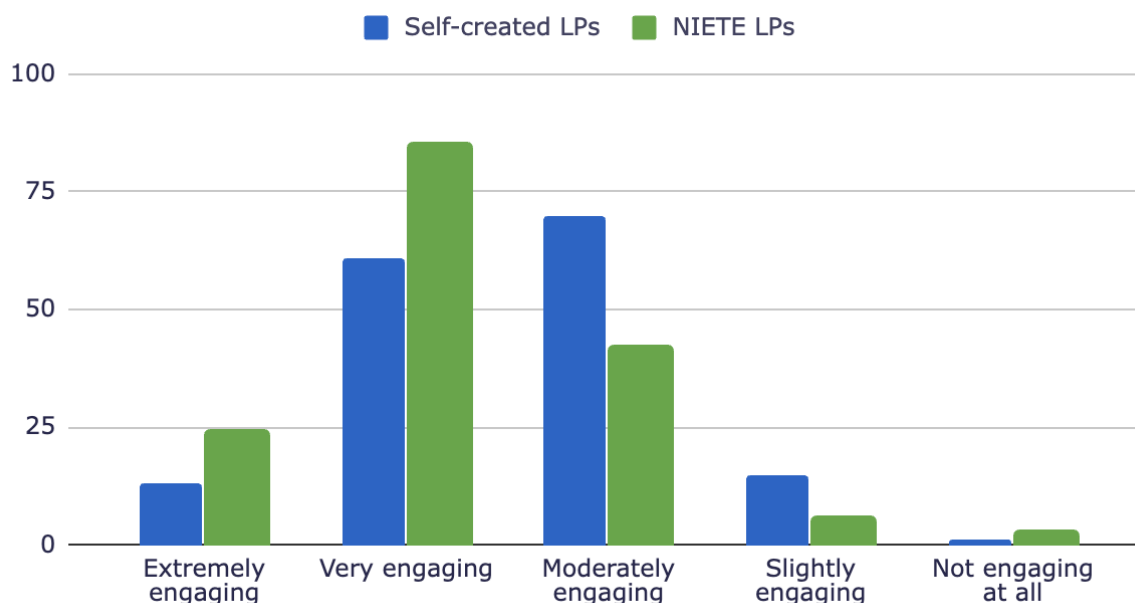
Qualitative data from the focus group discussions supports this result as teachers mentioned that student engagement improved with the help of the NIETE lesson plans. The structured activities in the NIETE lesson plans resulted in greater participation in class, and the frequent formative questions, group work, and hands-on activities helped generate and maintain student interest throughout the lesson.

> *"Kids also ask questions now. Kids who didn't use to participate also do now."* (FGD 1 participant)

Most of the teachers identified the activities in these lesson plans as significant drivers of student engagement, and this assertion is supported by data from the survey in which the teachers rated the activities in NIETE's lesson plans as better developed and easier to execute (see Figure 4.13).

Figure 4.13. *Survey responses: Development and execution of in-class activities*
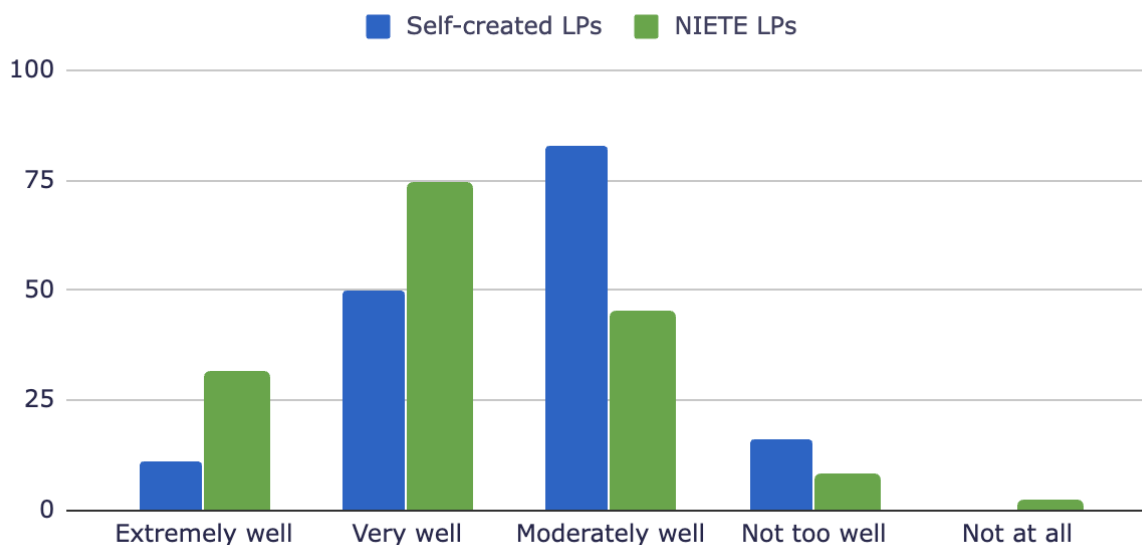
Qualitative data from the focus group discussions and the survey also reveals that some teachers find it difficult to manage the activities within the given time, as the activities can be very time-consuming in larger classes. While teachers appreciate the activities provided in the lesson plans, some pointed out the need for even more hands–on activities that use audiovisual aids and tools.

Expert review summary

The expert reviews found the lessons rich in engagement-facilitating activities. Frequent formative questions and pair or group activities ensured students actively participated in the lessons. In some plans, the activities were not strictly relevant to the objectives of the lessons and so seemed forced.

4.4. Expert reviews

Expert reviews were conducted to objectively evaluate the NIETE AI-generated lesson plans against a structured evaluation framework. Independent education experts assessed the lesson plans across multiple dimensions, including pedagogical alignment, clarity, assessment integration, and adaptability to student needs. This process aimed to validate the quality of the lesson plans by applying consistent criteria developed as part of the evaluation framework and rooted in established educational models such as the Gradual Release of Responsibility model, the 5E Instructional Model for science, and the Concrete-Pictorial-Abstract approach for mathematics. The expert evaluations complement the qualitative insights from teachers and key informants by offering a

systematic, evidence-based assessment of the lesson plans' design and instructional effectiveness. This section synthesises the findings from the expert reviews, highlighting areas of strength and opportunities for improvement in the lesson plans' structure and implementation.

The expert reviews were designed to provide quantitative scores for the lesson plans, using a scoring rubric (see Appendix), and qualitative trends across the identified themes for evaluation. A summary of the results of the scores each lesson plan received is shown below:

Table 4.1. *Rubric scores of the sampled lesson plans*

	Mathematics	**GK/Science**	**Urdu**	**English**
Grade 1	38	35	29	27
Grade 2	37	29	17	28
Grade 3	27	35	30	23
Grade 4	32	27	38	36
Grade 5	33	27	25	20
Average	33.4	30.6	27.8	26.8
Rating	**Good**	**Good**	**Satisfactory**	**Satisfactory**

Expert review data supplemented with data from the key informant interviews provides the following key insights about the NIETE lesson plans:

- They align well with the curriculum. Most lesson plans adhere to national curriculum guidelines and integrate subject-specific pedagogical approaches, such as Concrete Pictorial Abstract for mathematics.

- They provide structure and activities that are mostly lacking in lesson plans used in public or low-cost private schools across Pakistan, but can be improved when evaluated against established frameworks and international quality benchmarks.

- They deploy inconsistent differentiation strategies. While some plans provide scaffolding, many lack clear differentiation for struggling or advanced learners.

- There are gaps in conceptual depth and transitions. Lesson plans sometimes fail to build foundational understanding before

introducing more complex tasks, leading to abrupt transitions that may confuse students.

■ There is a mismatch between instruction and assessment. In most lesson plans, assessments test factual recall rather than the skills explicitly taught in the lesson.

■ There are frequent activities and good tools for student engagement, but across the sampled lesson plans, variety is lacking, feedback mechanisms are limited, and relevance to the learning objectives is sometimes weak.

The expert reviews against established frameworks and benchmarks shed light on the objective quality of the lesson plans. The overall 'Satisfactory' to 'Good' score of the lesson plans indicates their overall effectiveness but also points to the need for improvements, especially in assessment strategy and clarity of explanations. However, qualitative data from key informant interviews and the experts' opinions suggest the NIETE lesson plans are an improvement on the quality of the existing prevalent lesson plans and materials being used in public and low-cost private schools in Pakistan.

5. Conclusions and recommendations

This section synthesises the key insights derived from the mixed-methods analysis of the NIETE AI-generated lesson plans, combining qualitative data from focus group discussions and key informant interviews with quantitative findings from teacher surveys and expert evaluations. The conclusions highlight the strengths and areas for improvement in the lesson plans, particularly in their pedagogical quality, ease of use, and alignment with instructional frameworks.

Based on the findings, the recommendations emphasise the importance of incorporating greater flexibility in lesson plan design to accommodate diverse classroom contexts, unplanned disruptions, and varying student learning paces. Additionally, they advocate for promoting teacher agency in the implementation process and providing comprehensive professional development programmes to equip educators with the skills needed to use AI-generated materials effectively.

Recognising the need for sustainable and equitable scaling, the recommendations also highlight the importance of investing in digital infrastructure and fostering an iterative design approach informed by empirical evidence and stakeholder feedback.

As the NIETE programme expands to other regions, these insights offer practical guidance for policymakers, educational leaders, and technology developers to optimise the lesson plans and ensure they effectively support teaching and learning outcomes.

5.1. Synthesis of insights

The synthesis of insights combines the qualitative and quantitative findings to present a cohesive understanding of how the NIETE lesson plans function in actual classroom settings. This synthesis identifies recurring themes across data sources, emphasising areas where the lesson plans effectively support teaching and where further improvements are needed. There was no homogenous set of previously used lesson plans to carry out a comparative analysis, so the research focus was broadened to include the lesson plan development processes and delivery preparation efforts. By consolidating perspectives from teachers, independent experts, and key informants, the analysis provides a balanced reflection in comparing the AI-generated lesson plans to previously used material, the ability to enhance instructional delivery, and respond to the diverse needs

of students and educators, while maintaining a comparison with previously used material.

5.1.1. Comparative analysis

The evaluation of NIETE's AI-generated lesson plans against teacher-developed lesson plans revealed significant insights into their effectiveness, usability, and impact on instructional quality. Teachers widely acknowledged NIETE's strengths in providing a structured, standardised framework that ensured consistency across classrooms. However, the findings also highlight areas where greater flexibility and adaptability could enhance the lesson plans' practical utility. The insights presented here are drawn from quantitative survey data, focus group discussions, and key informant interviews, and form the basis for key takeaways and improvements.

The bar chart in Figure 5.1 below presents key survey findings that informed these conclusions, and illustrates a clear preference by teachers for NIETE's lesson plans across several domains, including differentiation, engagement, and in-class activities. NIETE's lesson plans were preferred for their structured framework and adherence to national guidelines, aligning with findings from the focus group discussions.

Figure 5.1. *Percentage of high teacher ratings (4 or 5 on a 5-point Likert scale) for LP quality comparison*
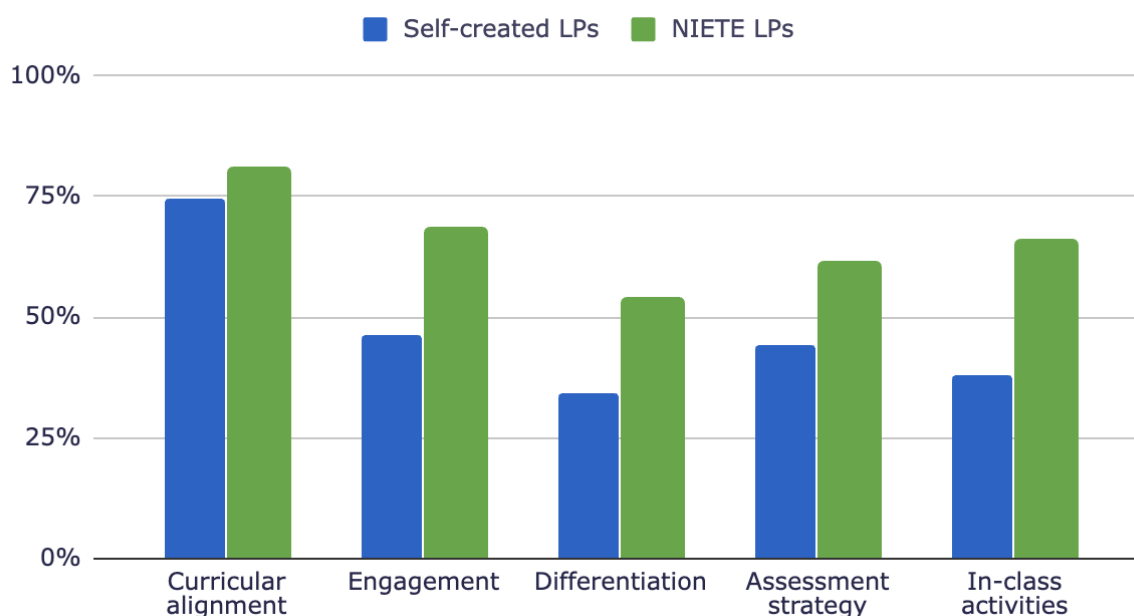
Findings from key informant interviews, focus group discussions, and the survey also critically revealed that, in many cases, teachers did not follow a structured pedagogical framework before they began using NIETE's lesson

plans. The existence, consistency, and quality of their self-created lesson plans varied significantly depending on factors such as school leadership oversight, teacher experience, and available resources. This lack of standardisation often led to inconsistencies in instructional delivery. Recognising this, the research focus was adapted from merely evaluating lesson plan content to examining differences in preparation effort, implementation ease, and overall instructional quality between the two approaches.

Teachers particularly valued the structured models employed in the NIETE lesson plans. While the standardisation provided much-needed consistency, some teachers expressed concerns about the rigidity of the lesson plans, feeling they limited their ability to personalise instruction. A participant in a focus group discussion emphasised, "These lessons should be made with a more pragmatic point of view, keeping in view time availability and the strength of the students in one class." Many teachers reported modifying NIETE's lesson plans to fit classroom constraints, such as adjusting for backlogs or accommodating students with different learning needs. Some teachers also found the length of lessons challenging. There were repeated suggestions to divide lesson plans into smaller, modular components that would allow for easier implementation. In addition, while NIETE introduced differentiation strategies often missing from the plans teachers created, teachers felt they needed more flexibility to make adjustments for varied student abilities. A survey respondent noted that NIETE's scaffolding techniques and engagement activities were particularly effective in supporting mixed-ability classrooms.

There was an overwhelming consensus that teachers found NIETE's lesson plans well-aligned with the curriculum, improving consistency across classrooms. In the survey responses, while teachers rated their own and NIETE's plans 'high' for curriculum alignment, they nonetheless rated NIETE plans higher than their own plans. Teachers also particularly valued the structured approach of NIETE's lesson plans, which reduced the time burden associated with lesson planning.

Engagement and in-class activities received significantly higher ratings in NIETE's lesson plans than teacher-created ones. Teachers found that pre-designed, interactive elements facilitated better student participation, with one stating: "In NIETE, after every paragraph, there are five questions, so kids understand better." However, teachers suggested that further improvements could be made by enhancing lesson interactivity through additional student-driven exercises, group discussions, and real-world applications.

One of NIETE's strongest advantages was its assessment strategy, as evident in Figure 5.1 above. Teachers found that the built-in formative assessments helped track student progress more effectively. As one participant noted: "Assessments are effective—we know where kids are at what stage." Compared to teacher-created lesson plans, which often lacked systematic evaluation methods, NIETE's structured approach reinforced student learning. However, teachers also reported scheduling challenges, particularly when NIETE's lesson progression did not align with routine school exams. One teacher remarked: "Routine exams didn't account for NIETE's schedule. Do we prep students for exams or continue NIETE lessons?" Aligning lesson pacing with school assessment cycles would enhance implementation without disrupting exam preparation.

While NIETE's approach was particularly beneficial for newer teachers, offering a clear instructional framework, more experienced teachers sometimes found its scripted nature restrictive. Many suggested additional training to help them modify lesson plans more effectively. Developing professional development workshops and providing modification guidelines would empower teachers to personalise their instruction while maintaining the benefits of a structured approach.

As highlighted in the literature review in Section 2, one of the key challenges for AI-generated content is finding the right balance between automation and teacher autonomy. This rapid evaluation reflects a similar challenge. Addressing teacher concerns around flexibility is a critical factor in ensuring seamless adoption. Creating a model that incorporates both standardisation and teacher agency is critical to maximising effectiveness. A blended approach—allowing teachers to customise AI-generated lesson plans while maintaining core instructional structures—could optimise effectiveness and usability. By addressing these considerations, NIETE's lesson plans can remain a valuable tool for enhancing instructional quality while accommodating the diverse needs of teachers and students.

5.1.2. Centralised lesson plans: A tool for structure and consistency

Findings from the key informant interviews and the evaluation of previously used instructional materials indicate that not all teachers consistently developed formal lesson plans for their lessons. Detailed discussions with the teachers during the focus group discussions revealed a consensus that developing lesson plans adds considerably to teachers' workload, especially for teachers who are also responsible for other

administrative tasks, and for teachers who teach more than two subjects, which is very common in primary grades.

> *"If a teacher has all subjects in one class, including Islamiat, and they also have to grade tests, so lesson plans is a burden [for them]. Lesson plans are also interrupted at times."*
> (FGD 1 participant)

Questions and discussion in the focus group discussions about best practices and necessary components of a lesson plan resulted in diverse, non-uniform responses that indicated a lack of comfortable familiarity with developing structured lesson plans. Furthermore, no standardised pedagogical model was guiding instructional delivery across schools. Teachers had the freedom and flexibility to develop their lesson plans according to their contexts and curricular requirements. In contrast, the NIETE lesson plans' integrated deployment of internationally recognised and evidence-based pedagogical models provides teachers with a structured approach to delivering lessons in the classroom, potentially benefiting learners with the improved learning outcomes promised by such established models.

In terms of additional resources for activities and assessments, teachers either relied on the activities provided in the textbooks or on their own judgment to design classroom activities, often requiring significant time for brainstorming and collaborative discussions to develop relevant learning experiences.

The assessment or evaluation of student progress or learning was left to summative assessment mechanisms and formative assessment questions were used spontaneously rather than being systematically integrated at key checkpoints, resulting in an unstructured approach to assessing student understanding. The NIETE lesson plans provided frequent and structured checkpoints to gauge student understanding that also resulted in greater participation and engagement.

The absence of a predefined lesson structure led to considerable variability in lesson quality across different schools, geographic areas, and under various administrative leaders. This flexibility, while allowing for some teacher autonomy, contributed to inconsistencies in instructional delivery and assessment practices overall. In contrast, the NIETE lesson plans provide a clear, structured framework grounded in established pedagogical models, including the Gradual Release of Responsibility model. Specific issues aside, all sampled teachers expressed strong appreciation for the NIETE lesson plans, recognising that their provision

significantly reduces the time and effort required for lesson preparation. This positive reception suggests that structured lesson plans can alleviate the planning burden while promoting a more consistent and pedagogically sound approach to teaching.

5.1.3. Flexibility and adaptability questions

This report outlines how a centralised repository of structured lesson plans is useful to bring structure and consistency across the 400+ schools under the Federal Directorate of Education's ambit. The downside to this approach is that it restricts the flexibility and adaptability of the lesson plans across the varying contexts of their delivery and implementation. According to data from the key informant interview with the Federal Directorate of Education administrator, the lesson plans are the same for every school and region and could be improved with further contextualisation.

> *"The scripted nature of these lesson plans make them difficult to contextualise according to their own (teachers and students) needs and environments — compare a school in F-6 (an area of elites) with Jhangi Syedan (a rural location)."*
> (Key Informant Interview 3)

The lesson plans and allotted times for the different sections are the same for class sizes ranging from 15 to 50 students. According to focus group discussion data, this meant that teachers could not complete activities in larger classes, or were unable to give struggling students extra time. Experienced teachers modified the NIETE lesson plans to cater for changing contextual requirements, while reflecting that weaker or less motivated teachers would compromise on learning to complete a lesson according to the lesson plans. This points to a gap in the utility and efficacy of the NIETE lesson plans between experienced/high-performing and inexperienced/underperforming teachers.

The adaptability and contextualisation issue is also raised in the expert reviews. For example, in an Urdu lesson featuring historical buildings, the lesson plan assumes all students have visited Lahore and seen its historical structures. The likelihood of this being the case or of the topic being contextually relevant might be high for schools in Punjab or even Islamabad, but it dramatically decreases in other provinces and regions. This means that contextualisation and adaptability will be significant questions to answer when scaling the use of AI-generated lesson plans in provinces like Balochistan.

5.1.4. Teacher perceptions and implementation fidelity

As mentioned in earlier sections, the scope of this research was mostly limited to teacher perceptions when comparing the NIETE lesson plans to previously used lesson plans and evaluating their quality. During the implementation of our study, the triangulation of teachers' usage data from the platform, their own responses on usage frequency in the survey, and qualitative data from the focus group discussions revealed contradictory, yet considerably insightful, findings. The contradictions in usage data and self-reported usage, combined with qualitative discussions in the focus group discussions, provide relevant insights into the challenges faced by Taleemabad and questions around implementation fidelity.

First, the data from the platform and responses in the key informant interviews reveal that teachers struggle to stay 'on schedule' with NIETE's lesson plans. According to NIETE's own criteria for teachers being 'on schedule', 'moderate users', and 'not engaged', usage data from their portal reveals the following statistics (presented in Figure 5.2).

Figure 5.2. *Usage data from the NIETE platform*

For **25%** or more of the school weeks

3.1%

of teachers were
'on schedule'

<0.1%

of teachers were
'moderate users'

89.7%

of teachers were
'not engaged'

For **70%** or more of the school weeks

0.3%

of teachers were
'on schedule'

0%

of teachers were
'moderate users'

57.3%

of teachers were
'not engaged'

The statistics in Figure 5.2 above represent 3,423 teachers who accessed and read at least three NIETE lesson plans.

This data reveals that teachers struggle to stay 'on schedule' according to NIETE's set timelines and plans. Focus group discussion data reveals consensus about the efficacy of the NIETE lesson plans and the ease they provide in terms of reduced preparation load for the teachers. Survey data reveals 69% of teachers showing greater satisfaction with NIETE lesson plans than with their own plans.

Figure 5.3. *Survey question: Compared to your previous lesson planning process, how satisfied are you with the NIETE lesson plans?*

- Slightly less satisfied — 9.0%
- Neutral — 21.3%
- Slightly more satisfied — 33.5%
- Much more satisfied — 35.5%
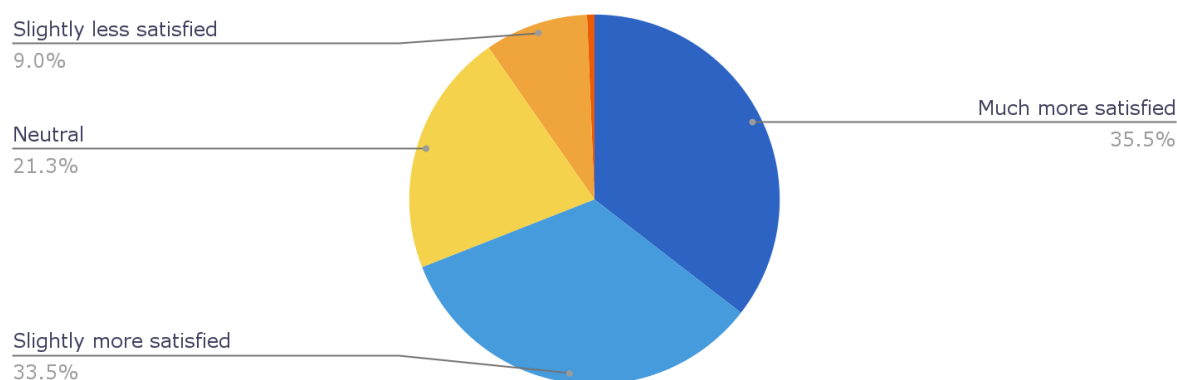
The question then becomes one about teachers' comfort levels and the skill required to implement these lesson plans. The adaptability question was explored earlier in Section 5.1.3 and the factors around adaptability of the lesson plans might be a factor affecting responses, but the survey question on levels of comfort with NIETE lesson plans also reveals that 66.3% of the respondents were either *extremely* or *very confident* about delivering the NIETE lesson plans.

Figure 5.4. *Survey question: How confident are you delivering NIETE's lesson plans compared to your previous method?*

- Slightly more confident — 6.3%
- Moderately confident — 25.6%
- Extremely confident — 18.8%
- Very confident — 47.5%
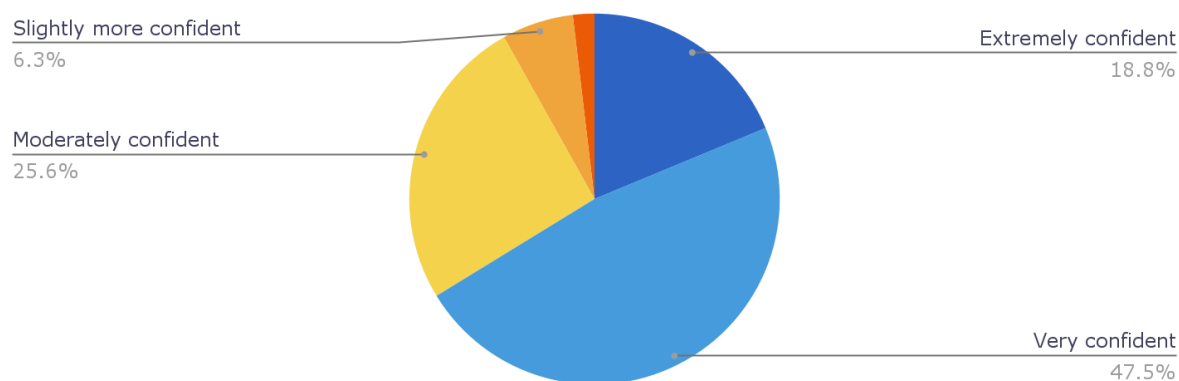
The comparative data from the surveys and focus group discussions shows that most teachers rated NIETE's lesson plans and all its components higher than they rated their own. This begs the question of why the adoption of lesson plans is so low and why teachers are not 'on schedule' with their use of this reportedly effective tool. A few insights from our analysis of the data might answer this question.

1. Teacher agency is being restricted

The messaging about the usage of the platform and its lesson plans directed at teachers and administrators is aggressive, and teachers

reported feeling their freedom and agency are being challenged and restricted. Focus group discussion data shows that teachers were very happy with the NIETE lesson plans but gave negative feedback about the mechanism for delivery that has been set. Teachers noted that when the programme started, they were all against it due to the tone of engagement, but with time, they found real utility in the lesson plans.

"Once we get used to it, it will be fine." (FGD 2 participant)

"Takes a while to adjust, I'd say give it a year."
(FGD 2 participant)

The modes and methods used for monitoring and evaluation also elicited negative emotions in the teachers. Frequent classroom observations are seen as putting teachers on the spot and as a tool for limiting teacher agency. The "red lines" used under teachers' names to indicate those who are not able to stay 'on schedule' is considered infantilising by teachers, leading to confusion and frustration with the system.

"Classroom observations stress us out." (FGD 2 participant)

"Teachers should be included in classroom observations."
(Survey response)

"It is sort of threatening for us, the red line, we even get messages."
(FGD 2 participant)

"I was in the red line but [unclear] about the reason."
(FGD 2 participant)

Experienced teachers also had reservations about the scripted lesson plans, as they felt they reduced their agency to lengthen or shorten lessons according to context or to add content based on their expertise and years of experience.

2. There is a lack of flexibility

The scripted and tightly structured lesson plans mean teachers do not have the flexibility to modify the lesson plans according to their contexts. One major insight derived from the data is that the perception of the lesson plans being too lengthy is pervasive. Data from all sources shows that these lesson plans are individually manageable within the given time

frame. But if we consider disruptions, interruptions, delays, unplanned holidays, and transition time within a lesson from one section to the next, and add them all up, the misalignments are considerable.

> *"The difference was coming in the schedule. One month's difference. Routine exams, NIETE didn't take time out for that. We cover Jan syllabus, and then tests for students. But NIETE lessons were still going on. So do we do our exams and prep or NIETE lessons? 1st term, 2nd term, 5 exams."*
> (FGD 1 participant)

A few teachers mentioned that if an explanation, activity, or assessment took longer than planned, they would make adjustments for that time in subsequent lessons and their plans. However, with NIETE, a loss in time snowballs across lessons and leads to bigger scheduling issues.

3. The mode of access and delivery is restrictive

The NIETE lesson plans are only available on the NIETE platform and, until recently, they could only be accessed online. After getting consistent teacher feedback, the NIETE team allowed teachers to download the lesson plans and save them offline on their smart devices. All the relevant training is only available online, and according to the sampled teachers, the guidelines given by the coaches are insufficient.

> *"Don't bind us that we have to complete the training. I'm interested in LPs, but if my phone doesn't work, what do I do? I train on the phone. My app wasn't opening. I tried till 2 AM at night."* (FGD 1 participant)

In both the survey and the FGDs, teachers requested the ability to access the lesson plans as PDFs. As a reading device, smartphones are comparatively more restrictive than printed paper, and scrolling on a phone during class times can be distracting for both students and teachers.

4. Gaps between utility for experienced and inexperienced teachers are emerging

As discussed earlier, the NIETE lesson plans and their accompanying mechanism for delivery and monitoring restrict teacher agency when delivering a lesson. The lesson plans limit experienced teachers who want to leverage their experience and expertise for varying contexts and situations. Teachers feel the plans force their work to be more mechanical.

For inexperienced teachers, the lesson plans are proving to be a useful tool in providing structure and content for their lessons, but they nonetheless restrict the teachers' growth and professional development.

Data from the key informant interviews shows that while teachers accept that the lesson plans are a good resource and help reduce their workloads, teachers feel they also limit their growth and learning, and can be a reason for lower job satisfaction as they are restrictive due to their scripted nature.

5.1.5. Shortcomings in the curriculum and textbooks are being carried forward

Generally, curricula and student learning objectives are developed taking into account the diverse contextual variations in implementation and learning environments. Textbooks then interpret and deliver the curricula based on more specific contextual requirements and target schools. Teachers in these schools then use the textbooks and other supporting materials to deliver the curriculum in the classrooms. Through this process, the curriculum sets the guidance and standards, and the textbooks provide contextualised material based on those guidelines and standards. Teachers then further contextualise the curriculum delivery based on their students' immediate requirements and needs.

Data from the key informant interviews reveals that most teachers are unfamiliar with the curriculum. They prefer to follow the textbooks; for most of them, "[the] textbook is the curriculum". It is also generally accepted that textbooks in Pakistan do not have the highest standard and include conceptual and other mistakes relatively often.

Analysis of the expert review reflects that mistakes or shortcomings in the textbooks are not filtered through the lesson planning process at NIETE and the tight links to the textbooks result in the same mistakes being repeated or duplicated in the lesson plans. Furthermore, the scripted nature of the lesson plans limits teachers' ability to make adjustments that previously allowed them to address errors in the textbooks. As a result, an opportunity to correct such inaccuracies and prevent them from reaching the students is now being missed.

5.2. Strategic recommendations

Based on the findings from the experts' review, focus group discussions, key informant interviews, and teacher surveys, this section outlines strategic recommendations to enhance the quality and implementation of the NIETE AI-generated lesson plans. These recommendations aim to strengthen the lesson plans' pedagogical effectiveness, improve their

adaptability across diverse educational contexts, and support sustainable scaling efforts as the programme expands to other provinces, starting with Balochistan. By addressing key areas such as lesson plan design, teacher support, and assessment integration, these recommendations provide actionable insights for policymakers, educational leaders, and technology developers. The goal is to ensure that AI-generated materials not only reduce teacher workload but also maintain high instructional standards and respond effectively to the contextual needs of both educators and students.

1. Incorporate flexibility in the design of the lesson plans

The scripted nature of the lesson plans limits teachers' ability to modify and adapt them based on their expertise and specific classroom needs. Given their centralised development, it is crucial for the lesson plans' design to allow flexibility. Flexibility is essential to accommodate a range of contextual factors, including varying classroom sizes (from fewer than 10 to over 50 students), unplanned disruptions, the diverse capabilities of educators (ranging from novice teachers to those with over 30 years of experience), and the differing learning paces of students. Allowing teachers to adjust the pacing and content delivery ensures they can provide additional time and support where needed, rather than adhering strictly to the time-based breakdown provided in the lesson plans.

2. Facilitate and promote teacher agency

The highly scripted nature of the lesson plans, combined with the way the programme is structured, has led to concerns among teachers about their autonomy in the classroom. Their sense of agency is further diminished by the performance tracking system, where they may be flagged as low performers (e.g., by red lines under their names) due to factors beyond their control, such as unplanned school closures or interruptions in class time.

To address this, creating space for teacher input and adaptation within the lesson plans is essential. Providing flexibility in lesson delivery, incorporating mechanisms for teachers to give feedback on lesson effectiveness, and ensuring that performance metrics account for contextual challenges will empower educators and enhance their role in the learning process.

3. Invest in infrastructure and digital literacy

Teachers' access to the lesson plans is limited, as the plans can only be accessed via the NIETE app. To improve teacher access, investment in the requisite infrastructure and teachers' digital literacy is paramount. While this study focuses on the quality of the lesson plans, teachers were eager to provide feedback on the connectivity issues they face when trying to access the lesson plans or the training. This feedback and other considerable evidence point to gaps in infrastructure as well as digital literacy and fluency for the teachers.

4. Provide continuous professional development and support best practices

Teachers' positive reception of the structured lesson plans, combined with their limited experience with scripted lesson plans, present a valuable opportunity to offer professional development support and establish best practices for implementation. This support would empower teachers to effectively adapt and utilise the centrally provided lesson plans to best meet the needs of their specific classroom contexts.

5. Continue to foster iterative design processes

The iterative approach adopted by the NIETE team has been encouraging and has led to tangible improvements in the lesson plans, such as including activity options based on available resources. These enhancements are reflected in the increasingly positive reception from teachers over time. Sustaining this iterative process and incorporating feedback from teachers across diverse contexts will ensure continuous improvement and greater adaptability of the lesson plans as they are scaled to the provinces.

This study provides critical insights into the quality and usability of the NIETE AI-generated lesson plans, offering evidence to guide their refinement and future implementation. The findings highlight the potential of AI-generated materials to standardise instructional quality, reduce teacher workload, and support more structured pedagogical approaches in public education. As the programme scales to additional provinces like Balochistan, these insights can inform strategic decisions to enhance the effectiveness and adaptability of AI-driven educational tools across diverse contexts. However, the study is limited in its ability to assess the direct impact of these lesson plans on student learning outcomes due to the absence of experimental or quasi-experimental methods and baseline data. Further research is needed to evaluate long-term

educational impacts, explore how AI-generated materials influence student performance, and identify best practices for sustainable implementation. Ongoing evaluation and evidence-driven adjustments will be crucial to ensuring that AI-generated lesson plans continue to meet the evolving needs of teachers and students, supporting broader educational goals in Pakistan and similar contexts.

References

This bibliography is available digitally in our evidence library at
https://docs.edtechhub.org/lib/XKIEQVIQ

Akeyampong, K., Andrabi, T., Banerjee, A., Banerji, R., Dynarski, S., Glennerster, R., Grantham-McGregor, S., Muralidharan, K., Piper, B., Ruto, S., Saavedra, J., Schmelkes, S., & Yoshikawa, H. (2023). *2023 Cost-Effective Approaches to Improve Global Learning: What does recent evidence tell us are "Smart Buys" for improving learning in low- and middle-income countries?* World Bank. https://thedocs.worldbank.org/en/doc/231d98251cf326922518be0cbe30 6fdc-0200022023/related/GEEAP-Report-Smart-Buys-2023-final.pdf. (details)

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives* (1st edition). (details)

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2024). *Generative AI Can Harm Learning* (SSRN Scholarly Paper No. 4895486). Social Science Research Network. https://doi.org/10.2139/ssrn.4895486. Available from https://papers.ssrn.com/abstract=4895486. (details)

Baytak, A. (2024). The content analysis of the lesson plans created by ChatGPT and Google Gemini. *Research in Social Sciences and Technology*, 9(1), 329–350. https://doi.org/10.46303/ressat.2024.19. Available from https://www.ressat.org/index.php/ressat/article/view/815. (details)

Bybee, R., Taylor, J. A., Gardner, A., & Van Scotter, P. (2006). The BSCS 5E Instructional Model: Origins and Effectiveness. *BSCS (Biological Sciences Curriculum Study)*. https://www.researchgate.net/publication/281412517_The_BSCS_5E_ins tructional_model_Origins_and_effectiveness. (details)

Bybee, R., Taylor, J. A., Gardner, A., & Van Scotter, P. (2006). The BSCS 5E Instructional Model: Origins and Effectiveness. *BSCS (Biological Sciences Curriculum Study)*. https://www.researchgate.net/publication/281412517_The_BSCS_5E_ins tructional_model_Origins_and_effectiveness. (details)

Education Endowment Foundation. (2024). *Can Using ChatGPT for Lesson Planning Cut Teacher Workload?* Education Endowment Foundation.

https://educationendowmentfoundation.org.uk/news/can-using-chatgpt-for-lesson-planning-cut-teacher-workload. (details)

Fan, H., Chen, G., Wang, X., & Peng, Z. (2024). LessonPlanner: Assisting novice teachers to prepare pedagogy-driven lesson plans with Large Language Models. *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, 1–20. https://doi.org/10.1145/3654777.3676390. Available from http://arxiv.org/abs/2408.01102. (details)

Fisher, D., & Frey, N. (2008). *Better Learning Through Structured Teaching: A Framework for the Gradual Release of Responsibility, 3rd Edition*. ASCD (Association for Supervision and Curriculum Development). https://www.ascd.org/books/better-learning-through-structured-teaching-a-framework-for-the-gradual-release-of-responsibility-3rd-edition?variant=121031. Available from https://www.ascd.org/books/better-learning-through-structured-teaching-a-framework-for-the-gradual-release-of-responsibility-3rd-edition?variant=121031. (details)

Frøsig, T. B., & Romero, M. (2024). *Teacher agency in the age of generative AI: Towards a framework of hybrid intelligence for learning design* (No. arXiv:2407.06655). arXiv. https://doi.org/10.48550/arXiv.2407.06655. Available from http://arxiv.org/abs/2407.06655. (details)

GOV.UK. (2024). *Teachers to get more trustworthy AI tech, helping them mark homework and save time* [UK Government]. https://www.gov.uk/government/news/teachers-to-get-more-trustworthy-ai-tech-as-generative-tools-learn-from-new-bank-of-lesson-plans-and-curriculums-helping-them-mark-homework-and-save. (details)

Gurl, T. J., Markinson, M. P., & Artzt, A. F. (2024). Using ChatGPT as a lesson planning assistant with preservice secondary mathematics teachers. *Digital Experiences in Mathematics Education*. https://doi.org/10.1007/s40751-024-00162-9. Available from https://link.springer.com/10.1007/s40751-024-00162-9. (details)

Harris, L. M., Archambault, L., & Shelton, C. C. (2023). Issues of quality on Teachers Pay Teachers: An exploration of best-selling U.S. history resources. *Journal of Research on Technology in Education*, *55*(4), 608–627. https://doi.org/10.1080/15391523.2021.2014373. Available from https://www.tandfonline.com/doi/full/10.1080/15391523.2021.2014373. (details)

Karpouzis, K., Pantazatos, D., Taouki, J., & Meli, K. (2024). *Tailoring education with GenAI: A new horizon in lesson planning* (No. arXiv:2403.12071). arXiv. https://doi.org/10.48550/arXiv.2403.12071. Available from http://arxiv.org/abs/2403.12071. (details)

Kaufman, J. H., Woo, A., Eagan, J., Lee, S., & Kassan, E. B. (2025). *Uneven Adoption of Artificial Intelligence Tools Among U.S. Teachers and Principals in the 2023–2024 School Year*. RAND Corporation. https://www.rand.org/pubs/research_reports/RRA134-25.html. (details)

Khurshid, S., Khurshid, S., & Toor, H. K. (2024). Learning for an uncertain future: Artificial intelligence a challenge for Pakistani education system in the era of digital transformation. *Qualitative Research Journal*, *ahead-of-print*. https://doi.org/10.1108/QRJ-02-2024-0045. Available from https://www.emerald.com/insight/content/doi/10.1108/qrj-02-2024-0045/full/html. (details)

Kong, B., Li, Z., Ratner, S., University of Oxford, Hemberg, E., AnyScale Learning For All, & Haque, A. K. M. S. (2024). *Artificial Intelligence for Education in Bangladesh: Insights from the First Small-Scale Artificial Intelligence Needs Assessment and Training for Primary Education Teachers*. Asian Development Bank. https://doi.org/10.22617/WPS240575-2. Available from https://www.adb.org/publications/artificial-intelligence-education-bangladesh. (details)

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, *41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2. Available from https://www.tandfonline.com/doi/full/10.1207/s15430421tip4104_2. (details)

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, *41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2. Available from https://www.tandfonline.com/doi/full/10.1207/s15430421tip4104_2. (details)

Meyer, A., Rose, D. H., & Gordon, D. (2014). *Universal Design for Learning: Theory and Practice*. CAST Professional Publishing, an imprint of CAST, Inc. (details)

Ministry of Education, Singapore. (2024). *Transforming Education through Technology: Masterplan 2030*.

https://www.moe.gov.sg/education-in-sg/educational-technology-journey/edtech-masterplan. (details)

Molina, E., Cobo, C., Pineda, J., & Rovner, H. (2024). *AI Revolution in Education: What You Need to Know* (In Digital Innovations in Education). World Bank. https://documents1.worldbank.org/curated/en/099734306182493324/pdf/IDU152823b13109c514ebd19c241a289470b6902.pdf. (details)

OECD. (2024). *The Potential Impact of Artificial Intelligence on Equity and Inclusion in Education* (OECD Artificial Intelligence Papers No. 23). https://www.oecd.org/en/publications/the-potential-impact-of-artificial-intelligence-on-equity-and-inclusion-in-education_15df715b-en.html. (details)

Pearson, P. D., & Gallagher, M. C. (1983). The instruction of reading comprehension. *Contemporary Educational Psychology*, *8*(3), 317–344. https://doi.org/10.1016/0361-476X(83)90019-X. Available from https://linkinghub.elsevier.com/retrieve/pii/0361476X8390019X. (details)

Strielkowski, W., Grebennikova, V., Lisovskiy, A., Rakhimova, G., & Vasileva, T. (2024). AI-driven adaptive learning for sustainable educational transformation. *Sustainable Development*, sd.3221. https://doi.org/10.1002/sd.3221. Available from https://onlinelibrary.wiley.com/doi/10.1002/sd.3221. (details)

UNESCO. (2024). *Global Education Monitoring Report 2024: Pacific: Technology in education – A tool on whose terms?* GEM Report UNESCO; Commonwealth of Learning. https://doi.org/10.54676/FILL2531. Available from https://unesdoc.unesco.org/ark:/48223/pf0000391211. (details)

UNESCO. (2024). *Use of AI in education: Deciding on the future we want*. https://www.unesco.org/en/articles/use-ai-education-deciding-future-we-want. (details)

Uygun, D. (2024). Teachers' perspectives on artificial intelligence in education. *Advances in Mobile Learning Educational Research*, *4*(1), 931–939. https://doi.org/10.25082/AMLER.2024.01.005. Available from https://www.syncsci.com/journal/AMLER/article/view/AMLER.2024.01.005. (details)

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, *252*, 124167. https://doi.org/10.1016/j.eswa.2024.124167. Available from

https://linkinghub.elsevier.com/retrieve/pii/S0957417424010339.
(details)

Wiggins, G. P., & McTighe, J. (2005). *Understanding by design* (Expanded
2nd ed). Association for Supervision and Curriculum Development.
(details)

Xu, Z. (2024). AI in education: Enhancing learning experiences and student
outcomes. *Applied and Computational Engineering*, *51*(1), 104–111.
https://doi.org/10.54254/2755-2721/51/20241187. Available from
https://www.ewadirect.com/proceedings/ace/article/view/10863.
(details)

Appendix

Lesson Plan Evaluation

Review of Curriculum Alignment, Instructional Design, and Assessment

Prepared for: EdTech Hub
Prepared by: Dr. Aamna Pasha

Contributions by: Nawaz Aslam
Science Technical Lead (NCC)

Table of Contents:

INTRODUCTION

This evaluation provides an independent expert review of AI-generated lesson plans used in the NIETE programme in Islamabad Capital Territory. The analysis applies a structured evaluation framework to assess the quality, relevance, and effectiveness of these lesson plans in meeting national curriculum standards and pedagogical expectations. Key areas of focus include curricular alignment, content accuracy, assessment strategies, and adaptability to diverse learners. Additionally, the evaluation considers the pedagogical models embedded within the lesson plans to determine their instructional coherence and effectiveness in real classroom settings.

EVALUATION FRAMEWORK

The lesson plans were evaluated using a structured approach based on the following components::

- Alignment with Curriculum and Standards
- Instructional Design and Pedagogical Coherence Content Quality and Depth
- Clarity, Depth, and Accuracy of Content
- Adaptability and Differentiation
- Engagement and Student-Centered Learning
- Ease of Use for Teachers
- Assessment Integration
- Accessibility and Differentiation

Each criterion was rated on a scale of **1 to 5**, with **5 being Excellent** and **1 being Poor**.

SUMMARY OF FINDINGS

The evaluation revealed some of the following overarching themes across the subjects:

- **Alignment with Curriculum:** Most lesson plans adhere to national curriculum guidelines and integrate subject-specific pedagogical approaches such as CPA for mathematics.
- **Use of Structured Instructional Models:** Most lessons follow well-established frameworks, including Gradual Release of Responsibility (GRR), which supports student progression from guided to independent learning.
- **Incorporation of Real-World Connections:** Some lessons effectively contextualize learning by using familiar cultural references or everyday examples.
- **Inconsistent Differentiation Strategies:** While some lessons provide scaffolding, many lack clear differentiation for struggling or advanced learners.
- **Gaps in Conceptual Depth and Transitions:** Lessons sometimes fail to build foundational understanding before introducing more complex tasks, leading to abrupt transitions that may confuse students.
- **Limited Student-Centered Learning:** Several lesson plans remain overly teacher-directed, with fewer opportunities for student-led inquiry, discussion, or collaborative activities.
- **Mismatch Between Instruction and Assessment:** In most, assessments test factual recall rather than the skills explicitly taught in the lesson.
- **Lack of well-structured formative assessments** that do not demonstrate sufficient rigor. Rather than incorporating a diverse range of assessment methods, most lesson plans frequently rely on repetitive, simplistic activities (e.g., thumbs-up/down responses). Additionally, many lesson plans do not include higher-order thinking tasks aligned with Bloom's Taxonomy, focusing primarily on rote memorization rather than promoting analysis, evaluation, or creative problem-solving.
- **Limited Feedback Mechanisms:** Some lessons rely primarily on verbal feedback, with few structured opportunities for self-assessment, peer review or teacher written review.

MATHEMATICS

Grade 1

Topic 4: Money

Lesson 3: Subtracting Pakistani money

Overall score - 38 (Excellent)

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide appropriate depth.	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear feedback mechanisms.	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Comments:

- Blends two distinct concepts: The SLO combines two learning objectives: Identifying and understanding the value of Pakistani money (conceptual knowledge) and using Pakistani money in subtraction problems (application in arithmetic). It is unclear whether the focus is on recognizing currency or solving subtraction problems involving money.

- Lack of Measurable Action: The phrase "understand the value" is vague and difficult to measure.

- Could provide additional support for struggling learners (for example counters or stones in low resource settings)

- Would be useful to add in a written feedback loop beyond the verbal component. Additionally, the use of a thumbs up/thumbs down response is a surface-level engagement strategy.

Grade 2

Topic 8: Geometry

Lesson 1: Introduction to horizontal and vertical lines

Overall score - 37 (Excellent)

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide appropriate depth.	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear feedback mechanisms.	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Comments:

- The lesson includes some scaffolding, which effectively supports student learning. However, it lacks sufficient differentiation to accommodate diverse learning needs. For struggling learners, incorporating concrete materials such as sticks or straws could provide hands-on support, reinforcing key concepts in a tangible manner. Additionally, using color coding for horizontal and vertical lines could aid in visual distinction. To further challenge advanced learners, the lesson could introduce diagonal lines, adding complexity and encouraging deeper spatial reasoning.

- The lesson demonstrates strong support for visual, kinesthetic, and auditory learners through its explanation process.

- The use of precise mathematical language could be improved to align with formal terminology and strengthen conceptual clarity. Additionally, the lesson could incorporate broader real-world connections, such as examples of buildings, structures, or everyday objects that illustrate the concepts being taught. This would reinforce the practical relevance of the lesson and deepen student engagement.

- The lesson primarily relies on verbal feedback, which, while valuable, may not provide sufficient opportunities for reflection and self-assessment. Integrating written feedback methods—such as peer reviews, self-checklists, or targeted questioning—would allow for a more structured evaluation of student progress.

- Furthermore, the homework on page 172 does not effectively reinforce classroom learning, as it does not require students to identify or differentiate key concepts covered in the lesson. The wrap-up exercise also falls short in ensuring full mastery of the content.

Grade 3

Topic 6: Temperature and Time

Lesson 4: Read Calendar

Overall score - 27 (Satisfactory)

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide appropriate depth.	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear feedback mechanisms.	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Comments:

- The lesson plan does not adequately distinguish between the solar and lunar calendars, despite the national curriculum explicitly requiring students to understand both. This omission may lead to gaps in comprehension, as students do not receive a clear explanation of how these calendar systems differ in structure, usage, and significance.

- The teacher's responses regarding the day on which a specific date falls are not aligned with the 2025 calendar, increasing the risk of errors in student learning.

- The lesson lacks sufficient opportunities for independent discovery and is highly teacher-led, with minimal space for student-driven exploration. Encouraging students to explore calendar patterns and relationships on their own would enhance pattern recognition skills and foster a deeper understanding of how dates follow cyclical structures. A more student-centered approach, such as allowing learners to choose and analyze dates themselves, would make the lesson more interactive.

Grade 4

Topic 5: Algebra

Lesson 5: Basic Input and Output Rules

Overall score: 32 (Good)

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide appropriate depth.	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear feedback mechanisms.	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Comments:

- The lesson lacks a clear transition between the multiplication and division rules, which may cause confusion for some students. Without an explicit connection, learners may struggle to understand how the two operations are related. A more structured approach—such as using comparative examples or bridging exercises—would help clarify the shift from multiplication to division.

- Additionally, the pictorial representation in the lesson is problematic. In the case of division, the sticks should be grouped separately to visually reinforce the concept of equal partitioning.

- The wrap-up activity lacks rigor, providing limited opportunities for students to apply their learning in a meaningful way. A stronger assessment component—such as word problems, real-world applications, or student-generated examples—would ensure deeper comprehension and retention.

- Furthermore, there is a misalignment between the classroom instruction and the textbook activity. While the class lesson focuses on identifying input and output, the textbook expects students to identify the rule governing the relationship between numbers.

- The lesson assumes familiarity with terms such as input and output, which may be unfamiliar to some students. Providing explicit language support—such as definitions, visual aids, or contextual examples—would help ensure all students can access the lesson content effectively.

Grade 5

Topic 3: Fractions

Lesson 2: Comparing unlike fractions

Overall score - 33 (Good)

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide appropriate depth.	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear feedback mechanisms.	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Comments:

- The lesson does not incorporate real objects or manipulatives to explain fractions, which may make it difficult for struggling learners to grasp the concept. Using tangible materials such as fraction strips, counters, or everyday objects (e.g., paper strips) would provide a concrete representation of fractions, making the lesson more accessible.

- The assessments are very basic and do not provide sufficient depth to evaluate student understanding.

ENGLISH

Grade 1

Topic 2: My Family

Lesson 2: : Understanding our families

Overall score - 27 (Satisfactory)

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide appropriate depth.	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear feedback mechanisms.	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Comments:

- The lesson plan appears too advanced for Grade 1 students, particularly as it is the second topic of the academic year. Expecting students to read and write full sentences at this stage may not align with their developmental level.

- The lesson lacks clarity in transitions, moving abruptly between instructional elements such as questions, diagrams, and flashcards.

- The lesson jumps between lower-order and higher-order questions without appropriate scaffolding. Students may struggle to engage with higher-order thinking tasks without progressive support, such as guided discussions, modeling, and gradual questioning techniques.

- The assessment component is not sufficiently robust to measure student understanding effectively.

- The family tree included in the lesson does not effectively support the intended learning objective. As a result, it functions as a meaningless resource rather than an aid to student understanding.

- The lesson does not acknowledge variations in students' routines, limiting its inclusivity. Since family structures and daily routines differ across households, the lesson should incorporate a broader range of experiences to ensure all students feel represented and engaged.

Grade 2

Topic 6: Be Honest

Lesson 1: : Reading

Overall score - 28 (Good)

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide appropriate depth.	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear feedback mechanisms.	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Comments:

- The lesson lacks strong student engagement strategies. It is entirely teacher-led and does not incorporate opportunities for students to express difficulties or ask about unfamiliar concepts.

- The assessment component is not sufficiently robust. The lesson does not provide clear methods for evaluating student understanding beyond passive observation.

- Finally, the lesson lacks certain inclusivity elements. The story features two boys and includes items that may not be relatable or accessible to all students. A more inclusive approach would involve diverse representation and adaptable content to ensure all learners feel represented and engaged.

Grade 3

Topic 10: Healthy Habits

Lesson 5: : Effective writing and grammar

Overall score - 23 (Satisfactory)

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide appropriate depth.	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lessons are mostly passive.	No active learning; entirely teacher-centered.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear feedback mechanisms.	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Comments:

- The reviewed lesson plan lacks clarity in distinguishing between different types of connectors. The national curriculum specifies that students should be able to use connectors for sequencing (e.g., *first, second, then*) and reasoning (e.g., *because, therefore*). However, this lesson conflates the two without providing explicit differentiation.

- Additionally, the lesson lacks coherence in its instructional flow. It abruptly shifts between comprehension of the story and writing procedural texts without a clear transition or scaffolding. While it attempts to integrate reading and grammar-related student learning outcomes (SLOs), the connection between them is unclear, making it difficult for students to grasp the intended learning objectives.

- The lesson does not adequately foster student engagement. There is no structured approach to eliciting student-generated examples before guiding them in writing procedural texts. Incorporating an interactive step where students provide examples before moving into structured writing would enhance comprehension and application.

Grade 4

Topic 8: Good Study Habits

Lesson 3: Mastering Alphabetical Order

Overall score - 36 (Excellent)

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide appropriate depth.	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and	Activities promote deep engagement, inquiry,	Some interactive elements but lacks	Somewhat engaging but lacks	Few interactive components; lessons are mostly passive.	No active learning; entirely teacher-centered.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
Student-Centered Learning	discussion, and real-world connections.	deeper engagement strategies.	strong student involvement.		
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear feedback mechanisms.	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Comments:

- The lesson plan does not fully align with the national curriculum standard for Grade 4, which requires students to recognize the alphabetical arrangement of words based on the first three letters. Instead, the lesson focuses only on the first two letters, making the activity somewhat simplistic for this age group.

Grade 5

Topic 4: Unforgettable moments of my life

Lesson 5: Dialogue practice: Importance of planting trees

Overall score - 20 (Needs Improvement)

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide appropriate depth.	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lessons are mostly passive.	No active learning; entirely teacher-centered.
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
					without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear feedback mechanisms.	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Comments:

- The SLOs are disconnected, making it unclear what the central focus of the lesson is. There is no logical link between the learning objectives, which creates confusion about the lesson's intended purpose.

- The lesson assumes that all students know what a tourist guide is or have experience with travel, which may not be the case. Without a prior discussion or an introduction to the concept, students with limited exposure to tourism may struggle to engage with the material.

- The lesson fails to define what a dialogue is before asking students to create one. Without structured support—such as examples, sentence starters, or guided practice—students may struggle with this task.

- The instructions lack clarity, making it difficult for students to understand what is expected of them. Additionally, the explanation on trees has no connection to the guided practice activity on dialogue writing, leading to a major disconnect in lesson flow. Later, in independent practice, students are asked to read out their dialogues as a role-play, yet the concept of role-play is never introduced or explained.

- The lesson lacks a clear assessment strategy to check for understanding of dialogue writing. The assessment tasks focus on basic lower-order thinking skills, particularly in the discussion on planting trees, without assessing the deeper skills related to dialogue construction

URDU

Grade 1

Topic 1: اعاده

Lesson 26: تشديد

Overall score - 29 (Good)

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide appropriate depth.	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear feedback mechanisms.	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Comments:

- The lesson moves too quickly into writing, without providing sufficient gradual support for students to build their understanding.

- The explanation of Tashdeed lacks conceptual clarity and depth. The lesson does not adequately explain why Tashdeed exists, its function in pronunciation and meaning, or provide sufficient examples to illustrate its importance. Additionally, the lack of clear progression in instruction

Grade 2

Topic 6: چڑیا اور چوبا

Lesson 2: چڑیا اور چوبا۔ قبل از پڑھائی

Overall score - 17 (Poor)

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide appropriate depth.	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear feedback mechanisms.	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Comments:

- The Student Learning Outcomes (SLOs) emphasize reading comprehension, yet the lesson itself is entirely focused on spelling and pronunciation of difficult words. This disconnect between what is intended and what is actually being taught weakens the lesson's effectiveness.

- The lesson lacks depth, as it does not go beyond simple word recognition and pronunciation. There is no meaningful engagement with the words in context, such as discussing their meanings, usage in sentences, or connections to the overall text.

- The difficult words segment does not incorporate any comprehension-based activities. Instead, the focus remains solely on reading the words aloud, missing an opportunity to develop students' understanding and application of these words in real-world or literary contexts.

- The assessment exercise at the end does not connect to the content taught in the lesson, making it ineffective in measuring student learning.

Grade 3

Topic 10: جس کا خواب تھا دل کش

Lesson 5: استفہامیہ اور اقراری جملے
*Note: There is an error in the app where the front page for Grade 3 identifies Topic 10 differently from how it is labeled within each topic

Overall score - 30 (Good)

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide appropriate depth.	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear feedback mechanisms.	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Comments:

- The independent practice activity focuses only on creating interrogative sentences, whereas the SLOs and the rest of the lesson emphasize differentiating between interrogative and declarative sentences and constructing both types. This creates a disconnect, as students are not given the opportunity to apply the full range of skills outlined in the lesson objectives.

- The assessment activity does not evaluate students' understanding of sentence types but instead tests factual recall from the text, which is unrelated to the lesson's core focus.

Grade 4

Topic 14: زیبا کے پڑوسی

Lesson 2: (زیبا کے پڑوسی (سوال،جواب

Overall score - 38 (Excellent)

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide appropriate depth.	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear feedback mechanisms.	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Comments:

- This lesson could benefit from additional student-led activities and differentiation for different learning abilities.

Grade 5

Topic 19: ایک قدیم شہر

Lesson 5: تاریخی عمارت کے متعلق مضمون نویسی

Overall score - 23 (Satisfactory)

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide appropriate depth.	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear feedback mechanisms.	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Comments:

- The SLOs emphasize writing about a historical structure, yet the lesson unexpectedly shifts to idioms. The independent practice focuses entirely on idioms, while the guided practice and explanation center on essay writing and its components. This lack of cohesion makes the lesson confusing and does not effectively support students in achieving the intended learning objectives.

- The lesson assumes that all students have visited Lahore and seen these historical structures, which may not be the case. To ensure inclusivity, the lesson should incorporate visual aids, descriptions, or alternative structures from different regions (or in this case, Islamabad) so that all students can engage meaningfully with the topic.

- The lesson does not provide sufficient scaffolding for essay writing. Instead of breaking down the process step by step, it jumps into the expectation that students can write a full essay. Incorporating a structured writing framework and guiding students through one section at a time (e.g., brainstorming, outlining, drafting, revising) would strengthen their

writing skills.

- There is no focus on revising and editing student work. Effective writing instruction should include a revision phase, where students receive feedback and refine their essays.

GENERAL KNOWLEDGE

Grade 1

Topic 8: میرا جسم

Lesson 2: ہماری حسیں

Score: **35 points (Excellent)**

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate,	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
	and provide appropriate depth.				
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy;	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
	clear feedback mechanisms.				
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

- Follows the textbooks exactly. Curricular gaps in the textbooks are showing here as well. The third SLO of the lesson required more examples of 'feelings' from our senses e.g. hot and cold, bitter and savoury, pungent and flowery etc.
- Limited explanations and opportunities to differentiate.
- Well structured activities will lead to greater engagement and ease of use for teachers but the volume of activities will leave teachers struggling with larger class sizes
- Alternates for gaps in resources have been provided in the lesson plans to aid adaptability.

Grade 2

Topic 10: پیشے
Lesson 1: مختلف پیشے

Score: **29 (Good)**

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate,	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
	and provide appropriate depth.				
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
	feedback mechanisms.				
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

- The activity is not well thought-out. The implementation will be difficult and the objective is unclear.
- The closing activity does not align well with the objectives.
- The lesson is very well structured and follows the 5E and IBL models.

Grade 3

Topic 8: وسائل اور ان کی اقسام

Lesson 3: وسائل کی اقسام

Score: **35 (Excellent)**

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate,	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
	and provide appropriate depth.		teacher interpretation.		
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
	feedback mechanisms.				
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

- Well-structured lesson plan with a strong opening.
- Lacks flexibility for teachers to engage more with different sections or formative questions
- Mechanical in implementation.

SCIENCE

Grade 4

Topic 5: Matter and its Characteristics

Lesson 2:   Understanding States of Matter

Score: **27 (Satisfactory)**

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate,	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
	and provide appropriate depth.		teacher interpretation.		
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
	feedback mechanisms.				
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

- The lesson plan jumps between fundamental concepts without bridging them or giving opportunities for further explanation.
- The opening activity is relevant but weakly integrated with the learning objectives.
- Depth in explanations is limited.

Grade 5

Topic 5: Light and Sound

Lesson 5: Speed of Sound in Materials

Score: **27 (Satisfactory)**

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate,	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
	and provide appropriate depth.		teacher interpretation.		
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
	feedback mechanisms.				
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

- Conceptual mistake in the Lesson Summary: "… how it travels faster in solids, liquids, and gases."
- Words like 'propagation' and 'material medium' are suddenly introduced without explanation.
- Exact same examples used as in the textbook, limiting better understanding
- Relies too heavily on the textbook for 'independent practice'.

Evaluation Framework for NIETE's AI-Generated Lesson Plans

This framework provides an independent expert with a structured approach to evaluating the quality of AI-generated lesson plans used in the NIETE programme in Islamabad Capital Territory. The evaluation will focus on relevance, curricular alignment, content quality, assessment or evaluation strategy, and adaptability, considering the pedagogical models embedded within the lesson plans.

Framework Components

1. Alignment with Curriculum and Standards

Indicators:

- Lesson objectives clearly align with the national curriculum.
- Content is mapped accurately to prescribed textbooks and learning standards.
- Lessons adhere to subject-specific pedagogical approaches (e.g., CPA, 5E, IBL).

2. Instructional Design and Pedagogical Coherence

Indicators:

- Follows the Gradual Release of Responsibility (GRR) model effectively.
    - "I Do" or "Explanation": Provides clear explanations and modeling.
    - "We Do" or "Guided Practice": Includes guided practice opportunities.
    - "You Do" or "Independent Practice": Facilitates independent student practice.
- Science lessons integrate 5E & Inquiry-Based Learning (IBL) elements appropriately.
    - Students are "Engaged" through relevant examples or "Inquiries"
    - The lessons progress along the lines of the frameworks provided by these models while adhering to the overall GRR design.
- Math lessons incorporate Concrete-Pictorial-Abstract (CPA) progression effectively.

3. Clarity, Depth, and Accuracy of Content

Indicators:

- Explanations are clear and logically structured.
- Key concepts are well-articulated with appropriate depth.
- Examples and illustrations enhance understanding.
- Content is factually accurate and free from errors.

4. Adaptability and Differentiation

Indicators:

- Provides strategies for different learning paces and student needs.
- Includes scaffolding for struggling learners.
- Allows for extension activities for advanced learners.
- Accommodates diverse classroom contexts (e.g., rural vs. urban settings).

5. Engagement and Student-Centered Learning

Indicators:

- Lesson activities encourage active participation.
- Use of real-world examples and contextualized scenarios.
- Opportunities for student discussion and collaboration.
- Inquiry-driven tasks in science and problem-solving in math.

6. Ease of Use for Teachers

Indicators:

- Lesson structure is intuitive and easy to follow.
- Clear instructions and guidance for implementation.
- Logical sequencing of activities.
- Does not create an additional burden on teachers.

7. Assessment Integration

Indicators:

- Includes well-structured formative assessment opportunities.
- Questions align with Bloom's Taxonomy and other assessment frameworks.
- Provides feedback mechanisms to guide student progress.
- Provides support for summative assessment to the teachers.

8. Accessibility and Inclusivity

Indicators:

- Uses inclusive language and examples.
- Avoids cultural or gender biases.
- Supports students with diverse learning needs.

Evaluation Methodology

This structured framework ensures a comprehensive, evidence-based evaluation of the AI-generated lesson plans while aligning with best practices in instructional design. The variety of models being integrated in the lesson plans presents a challenge in synthesising

their individual components to align with the overall goal of improvement of learning outcomes. Similarly, this customised and adaptive use of the different models requires a holistic overview to evaluate the greater impact while also considering specific implementation of the different components of the integrated pedagogical models. The following rubric provides a tool for independent experts/evaluators to assess the quality of these lesson plans.

Rubric for Expert Review

This rubric provides a structured method for independent experts to evaluate AI-generated lesson plans based on key quality dimensions. Each criterion is rated on a **5-point scale**, with descriptors outlining expectations at each level.

Criteria	5 - Excellent	4 - Good	3 - Satisfactory	2 - Needs Improvement	1 - Poor
1. Alignment with Curriculum and Standards	Fully aligns with the national curriculum and textbooks, ensuring coherence and accuracy.	Mostly aligned with minor deviations that do not impact learning.	Some alignment, but noticeable gaps in coherence.	Significant misalignment that may confuse teachers or students.	Major misalignment; does not reflect curriculum standards.
2. Instructional Design and Pedagogical Coherence	Fully adheres to GRR model; structured transitions between "I Do, We Do, You Do"; Science lessons integrate 5E/IBL; Math lessons use CPA effectively.	Mostly follows pedagogical models with minor inconsistencies.	Inconsistently applies models; lacks coherence in instructional flow.	Weak adherence to pedagogical models; lacks clear structure.	Does not follow GRR or subject-specific frameworks.
3. Clarity, Depth, and Accuracy of Content	Concepts are clearly explained, well-structured, factually accurate, and provide	Minor clarity or depth issues but still effective.	Some sections lack clarity, depth, or accuracy, requiring teacher interpretation.	Noticeable issues with clarity, accuracy, or depth that hinder learning.	Major flaws in clarity, accuracy, or depth, making the lesson unusable.

	appropriate depth.				
4. Adaptability and Differentiation	Provides strong scaffolding, differentiation strategies, and extension activities for all learners.	Includes some differentiation but lacks comprehensive scaffolding.	Limited adaptability; mostly designed for average learners.	Minimal differentiation; may not support diverse learners well.	No differentiation; does not consider diverse student needs.
5. Engagement and Student-Centered Learning	Activities promote deep engagement, inquiry, discussion, and real-world connections.	Some interactive elements but lacks deeper engagement strategies.	Somewhat engaging but lacks strong student involvement.	Few interactive components; lesson is mostly passive.	No active learning; entirely teacher-centered.
6. Ease of Use for Teachers	Clear, well-structured, and easy to implement with step-by-step guidance.	Mostly easy to follow, with minor areas for improvement.	Somewhat difficult to follow; requires extra effort from teachers.	Requires substantial teacher effort to interpret and implement.	Confusing and hard to implement without additional preparation.
7. Assessment Integration	Includes well-structured formative assessments aligned with Bloom's Taxonomy; clear	Assessments are present but could be better aligned or structured.	Some assessment integration but lacks rigor or clarity.	Limited or weak assessment strategies.	No meaningful assessment components included.

	feedback mechanisms.				
8. Accessibility and Inclusivity	Content is culturally appropriate, inclusive, and free from bias; accommodates diverse learners.	Mostly inclusive, with minor areas for improvement.	Some biases or lack of accommodation for diverse students.	Limited inclusivity; may not meet the needs of all students.	Content is exclusionary, biased, or inaccessible.

Scoring Guide

- **40-35 points (Excellent):** Highly effective lesson plan with strong alignment, clarity, adaptability, and engagement.
- **34-28 points (Good):** Generally effective but with minor gaps in content depth, differentiation, or engagement.
- **27-21 points (Satisfactory):** Adequate but with noticeable gaps that may require teacher adaptation.
- **20-14 points (Needs Improvement):** Significant flaws in clarity, alignment, or implementation, requiring revision.
- **Below 14 points (Poor):** Major flaws that limit usability and effectiveness; substantial revision needed.

Justification for the Evaluation Framework

This rubric ensures a structured, evidence-based review of AI-generated lesson plans, guiding expert evaluators to provide consistent and actionable feedback. Alignment with the curriculum is essential for ensuring coherence with national education standards. Instructional design elements, including the Gradual Release of Responsibility (GRR), 5E, Inquiry-Based Learning (IBL), and Concrete-Pictorial-Abstract (CPA) models, are used to assess the pedagogical effectiveness of the lesson plans. Clarity, depth, and accuracy of content are necessary to ensure that teachers receive well-structured and factually correct materials. Adaptability and differentiation are included to assess whether lesson plans can accommodate diverse learning needs, which is particularly crucial in heterogeneous classrooms. Engagement and student-centered learning principles are assessed to determine the effectiveness of interactive and inquiry-driven activities. Ease of use for teachers is a critical component, as overly complex lesson plans may lead to implementation challenges. Assessment integration ensures that formative assessment strategies are embedded to support student progress tracking. Lastly, accessibility and inclusivity are evaluated to confirm that content is free from biases and supports learners from diverse backgrounds.

References

This evaluation framework is informed by established instructional design principles and research, including:

- Pearson & Gallagher (1983) on the Gradual Release of Responsibility (GRR) model.

- Fisher & Frey (2008), Better Learning Through Structured Teaching by (ASCD).

- Bybee et al. (2006) on the 5E Instructional Model.

- Mao, Y. (2023) Issues and Strategies in Inquiry-Based Learning Evaluation. *Open Journal of Social Sciences*, 11, 422-440. doi: 10.4236/jss.2023.114030

- Putri, H., Suwangsih, E., Rahayu, P., Nikawanti, G., Enzelina, E., & Wahyudy, M. (2020). Influence of Concrete-Pictorial-Abstract (CPA) Approach on the Enhancement of Primary School Students' Mathematical Reasoning Ability. Mimbar Sekolah Dasar, 7(1), 119-132. doi:https://doi.org/10.17509/mimbar-sd.v7i1.22574.

- Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, *41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

- Universal Design for Learning (CAST, 2011) for accessibility and inclusivity considerations.

This structured rubric, backed by established research, provides a rigorous yet practical method for evaluating AI-generated lesson plans.