PAKAR Pendidikan

p-ISSN: 1693-2226 ; e-ISSN: 2303-2219
Vol. 24, No. 1, January 2026
Page. 1-18
http://pakar.pkm.unp.ac.id/

Integration of Google Sheets with Deep Learning-based Evaluation Methods to Measure the Analytical Skills of PPG Students in Creating HOTS Questions

Integrasi *Google Sheet* dengan Metode Evaluasi berbasis *Deep Learning* untuk Mengukur Kemampuan Analitis Mahasiswa PPG dalam Pembuatan Soal HOTS

https://doi.org/10.24036/pakar.v24i1.914

Adip Wahyudi[1*], Slamet Arifin[1], Evi Fitriana[1], Feri Fahrian Maulana[2], Dynda Prista[1],
Muhammad Aris Firmansyah[3], Radya Wirawan Nurjati[4]
[1] Universitas Negeri Malang, Malang, Indonesia
[2] Universitas Brawijaya, Malang, Indonesia
[3] PT. Alfahuma Rekayasa Teknologi, Malang, Indonesia
[4] National Yang Ming Chiao Tung University, Hsinchu City, Taiwan
[*]E-mail: adip.wahyudi.fis@um.ac.id

Abstract

21st-century education requires the use of creative interactive tools to make learning outcome evaluation more effective. The main challenge is low student participation due to conventional assessment practices. The Google Sheet evaluation tool that integrates deep learning offers a solution to the problem of learning evaluation. This study focuses on integrating an evaluation tool into Google Sheets using deep learning in the Learning Evaluation course in the PPG program and on testing its impact on learning outcomes. The research method used R&D with the ADDIE model, with data collected through a pretest-posttest design across two groups: the experimental group that used deep learning-based Google Sheets evaluation tools and the control group that used a conventional approach. After normality and homogeneity tests confirmed that the statistical prerequisites were met, the analysis continued with a paired t-test. The results showed significance (p < 0.05), indicating that the deep learning-based Google Sheets evaluation tool was effective in improving learning outcomes, while also encouraging student motivation and engagement. In conclusion, the deep learning-based Google Sheets evaluation tool is suitable for implementation in the PPG program, with opportunities for further development through the enrichment of collaboration features and adaptive evaluation to strengthen the appeal and impact of learning.

Keywords: *Google sheets, deep learning, evaluation tools, teacher professional education.*

Abstrak

Pendidikan abad 21 menuntut pemanfaatan alat interaktif yang kreatif agar evaluasi hasil belajar lebih efektif. Tantangan utamanya adalah rendahnya partisipasi mahasiswa akibat praktik penilaian yang masih konvensional. Alat evaluasi *google sheet* yang integrasikan *deep learning* menawarkan solusi untuk permasalahan evaluasi pembelajaran. Penelitian berfokus mengintegrasikan alat evaluasi dengan *google sheet* berbasis *deep learning* pada mata kuliah evaluasi pembelajaran di program PPG serta menguji dampaknya terhadap peningkatan hasil belajar. Metode penelitian menggunakan R&D dengan model ADDIE dengan data dikumpulkan lewat desain *pretest posttest* pada dua kelompok, yakni kelompok eksperimen yang menggunakan alat evaluasi *google sheets* berbasis *deep learning* dan kelompok kontrol yang memakai pendekatan konvensional. Setelah uji normalitas dan homogenitas memastikan prasyarat statistik terpenuhi, analisis dilanjutkan menggunakan uji t berpasangan. Hasilnya menunjukkan signifikansi (p < 0,05), menandakan alat evaluasi *google sheet* berbasis *deep learning* efektif meningkatkan capaian belajar, sekaligus mendorong motivasi dan keterlibatan mahasiswa. Kesimpulannya alat evaluasi *google sheets* berbasis *deep learning* layak diimplementasikan pada program PPG, dengan peluang pengembangan berikutnya melalui pengayaan fitur kolaborasi dan evaluasi adaptif untuk memperkuat daya tarik dan dampak pembelajaran.

Kata Kunci: *Google sheet, deep learning*, alat evaluasi, pendidikan profesi guru.

1. Introduction

The global education paradigm has undergone a profound transformation. Many professions in the future will require advanced analytical skills and creativity, prompting countries such as Finland and Singapore to integrate Higher Order Thinking Skills into their education curricula (Dos Santos et al., 2025). This change is not merely a trend but a strategic necessity for developing competitive human resources (Kacmaz & Dubé, 2025; Maulana et al., 2023). In Indonesia, implementing HOTS faces various challenges. Research conducted by Khoiriyah & Shaleh (2025) shows that only a few schools can implement the Higher Order Thinking Skills curriculum effectively, while the rest still rely on conventional learning methods, especially in 3T (Underdeveloped, Frontier, and Outermost) areas. The imbalance between progressive policies and the actual state of educational infrastructure has led to systemic inefficiencies (Mir et al., 2024).

Education in the 21st century requires the development of Higher Order Thinking Skills (HOTS) as the foundation of learning. Although the Indonesian curriculum has adopted a 21st century learning approach, practice in the field is still hampered by the dominance of evaluation instruments that only measure Lower Order Thinking Skills (LOTS) (Nasbi, 2017). Teachers, as the main agents of education, play a vital role, so their competence needs to be honed through Teacher Professional Education programs (Saputra & Nofrion, 2022). The main focus of this competency development lies in analytical skills, which are conceptually defined as C4 (Analyzing) cognitive abilities according to Bloom's revised taxonomy, including the skills of distinguishing, organizing, and attributing elements of information to construct complex test stimuli. However, there is currently a significant gap between these competency requirements and the available evaluation infrastructure. PPG students are required to reconstruct material into high-level questions, but the evaluation instruments used remain conventional and manual, failing to objectively assess the semantic structure and cognitive level of the questions. As a result, the validity of HOTS question construction is not well verified.

Based on observations in the Geography PPG Program at Universitas Negeri Malang (UM), empirical findings indicate that 65% of 40 students in the program have difficulty formulating analysis and evaluation questions in accordance with Bloom's revised taxonomy. Classroom observations show that 70% of the 40 students' questions remain at the C1-C3 level, indicating a lack of contextualized question structure. Interviews with PPG lecturers reveal that the main problem is the absence of instruments capable of assessing question quality while providing objective feedback. This leads to differences in the quality of the questions produced. Google Sheets has been used to compile questions due to its ease of organizing items and basic statistical analysis (Surahman et al., 2024). However, the previous instrument, which used Google Sheets, lacked features to automatically evaluate the cognitive complexity of questions. In fact, deep learning technology has already been proven in research (Yudistira & Kom, 2024) to be capable of classifying text by cognitive difficulty level using a Natural Language Processing model. The integration of these two tools has the potential to provide a holistic solution, but no research has implemented it in a structured manner in the PPG context.

Deep learning can process structured and unstructured data, such as text, images, and audio, through multi-level representation learning, where each layer of the network transforms the input into increasingly abstract and meaningful representations (Nasser et al., 2025). This capability enables extensive data collection and algorithms that assign different weights to each element of text. Thus, when used in education, Deep learning has demonstrated its effectiveness in tasks such as classifying the cognitive difficulty level of questions, generating higher-order thinking skills questions, and performing automatic feedback analysis (Sumy et al., 2023; Vanhees et al., 2025).

As evidenced by research (Kasanah & Pratama, 2024), which achieved 92% accuracy in categorizing questions according to Bloom's revised taxonomy.

Google Sheets functions as more than just a spreadsheet application; it serves as an intelligent evaluation interface through the use of Google Apps Script customization features (Parra et al., 2021). This feature is a strategic key that enables the integration of Deep Learning into student worksheets. In the context of the Geography Teacher Education Program (PPG) at UM, this integration transforms conventional spreadsheets into an interactive analytical ecosystem. Students not only enter the question text, but the system also automatically processes it to diagnose the cognitive level in real time. This instant feedback mechanism directly supports students' analytical skills in formulating complex questions, making Google Sheets an effective pedagogical tool for validating evaluation instruments (Rebman et al., 2023).

Research has made substantial progress in using Deep Learning for educational assessment, as exemplified by a transformer model that achieved 85% accuracy in categorizing Higher Order Thinking Skills problems (Chen & Han, 2025). However, these programs operate independently and are not integrated with established educator platforms, such as Excel, so they cannot be used for teacher training. Research conducted by Wahyudi et al. (2025) in the PPG Program at Universitas Negeri Malang revealed that 80% of students prefer to use Excel to compile questions. However, this application does not have an automatic cognitive analysis function, and validators must check it offline. Research by Aniek et al. (2022) found that the limitations of traditional tools in providing objective assessments of the quality of Higher Order Thinking Skills questions create a gap between the demands for developing students' analytical skills and current evaluation methodologies. This problem is exacerbated by the prevalence of C1-C3 level questions (LOTS) created by students, as revealed in field observations.

Based on previous research, the researchers filled the research gap by integrating Google Sheets with a bidirectional LSTM that analyzes the linguistic structure of questions, classifies cognitive levels according to Bloom's taxonomy, and offers recommendations to improve higher-order thinking skills. This integration seeks to align technological capabilities with educational needs while improving the objectivity and consistency of the quality of questions prospective educators generate for students in geography lessons.

2. Literatur Review

2.1. HOTS and the Learning Evaluation Paradigm

This study is based on the Constructivist theory, which views question construction not merely as an administrative task but as a cognitive process in which students construct meaning through authentic experiences (Sugrah, 2020). The main variable measured in this study, namely analytical ability, is based on Anderson & Krathwohl's Taxonomy. The main variable in Bloom's taxonomy has become the standard for content validity, emphasizing the higher-order cognitive processes of analyze, evaluate, and create (Kasanah & Pratama, 2024).

To support the development of these abilities, this study adopts the Assessment for Learning principle. Evaluation does not only serve as a final measurement tool but also as a continuous feedback mechanism that drives improvements in question quality (Misnawati et al., 2025). The integration of this theory confirms that improvements in PPG students' abilities can occur only if the system provides interventions in the form of specific suggestions for improvement on the draft questions they have compiled.

2.2. Deep Learning in Automated Assesment

The application of technology in this study is based on the latest literature on Deep Learning and Natural Language Processing. Research (Akdemir & Barışçı, 2024) shows the capacity of Deep Learning models to capture semantic patterns and natural language reasoning structures, which are relevant for detecting whether an item meets HOTS criteria or only LOTS. Specifically, the Deep Learning architecture in this system serves as an intelligent validator, replacing the slow, manual validation process. According to Gardner et al. (2021), automated systems offer advantages in assessment consistency. The inference pipeline, from text preprocessing to result visualization, produces actionable diagnostic analytics (Dagal et al., 2025). This converts raw scores into qualitative feedback that guides students in revising their stimuli and answer options.

In conclusion, Deep Learning in this study serves as an intelligent validator that automatically distinguishes the quality of HOTS and LOTS questions through language pattern analysis. Its existence overcomes the weaknesses of manual validation by offering greater speed and consistency in assessment. More than just giving scores, this system provides concrete feedback for improvement, directly guiding students to refine the quality of the questions they create.

2.3. Technology Integration and Cognitive Load

The selection of Google Sheets as the main interface is justified by Cognitive Load Theory (CLT). Research (Gkintoni et al., 2025) emphasizes the importance of minimizing extraneous cognitive load so that users' mental energy is focused on germane load. Using a familiar spreadsheet platform minimizes technical barriers for students, allowing them to focus on the substance of question development. The acceptance of this technology was also analyzed through the “Technological Pedagogical Content Knowledge” and “Technology Acceptance Model” frameworks. Technological Pedagogical Content Knowledge explains the synergy among geography content, pedagogical evaluation, and AI technology, while the Technology Acceptance Model predicts that students' perceptions of Google Sheets' ease of use will increase their adoption of the system (Huang et al., 2025). This interactive format has been empirically proven to increase engagement and reduce boredom in complex learning processes.

In conclusion, using Google Sheets aims to free students from the mental energy required to learn complex new applications, allowing them to focus fully on the substance of question preparation. Because this platform is already very familiar in their daily lives, the advanced technology embedded in it feels light and easy to accept, with few significant technical barriers. This familiar approach ultimately makes the learning process more lively, less boring, and able to increase student participation.

2.4. Conceptual Framework

Based on the synthesis of the above theories, this study's conceptual framework links problems, technological solutions, and expected outcomes. The process begins with student input by creating draft questions that are often still at the LOTS level. Through a system intervention that integrates Deep Learning algorithms into the Google Sheet interface, the system performs real-time semantic structure analysis. The system's output is not only classification labels, but also constructive feedback that triggers students to make independent revisions. This iterative cycle gradually builds students' understanding, ultimately leading to an increase in the study's dependent variable: analytical skills in developing HOTS-based evaluation instruments.

3. Research Method

The Google Sheets integration system with a deep learning-based evaluation method was developed using a user-centered ADDIE approach in this product. The researchers chose this approach because it can overcome complex challenges by putting themselves in the position of users, namely PPG students and teachers, iterating quickly on prototypes, and producing creative solutions that can be applied in geography classes. The five adaptive stages of the ADDIE process, as described by Maxnun et al. (2024), are analysis, design, development, implementation, and evaluation, as illustrated in Figure 1.

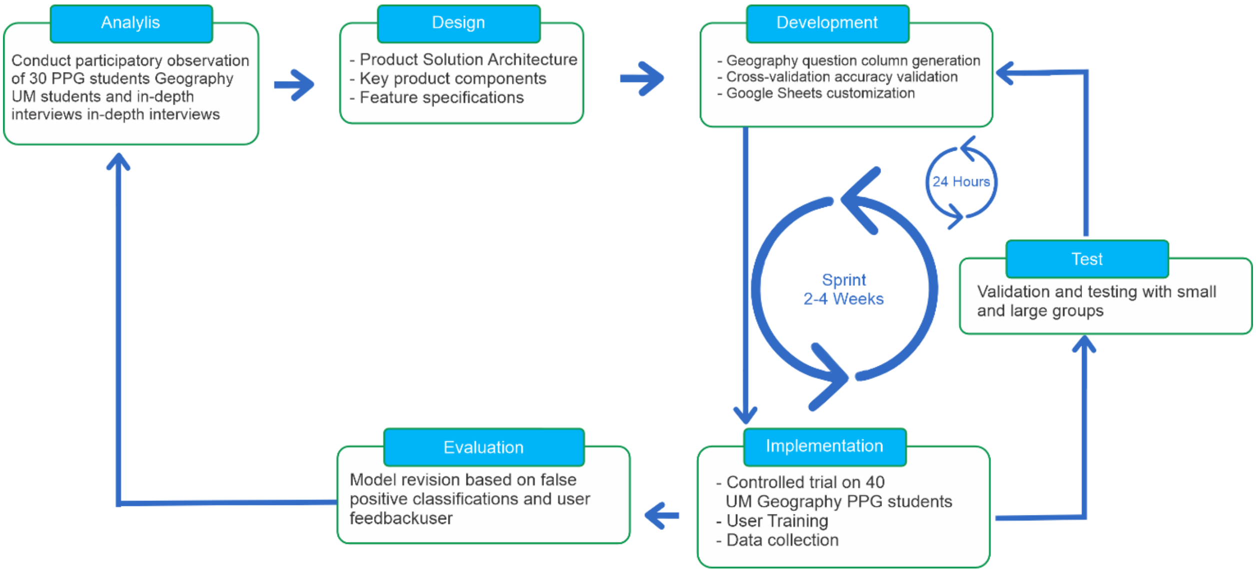

Analylis
Conduct participatory observation of 30 PPG students Geography UM students and in-depth interviews in-depth interviews
Design
- Product Solution Architecture
- Key product components
- Feature specifications
Development
- Geography question column generation
- Cross-validation accuracy validation
- Google Sheets customization
24 Hours
Sprint
2-4 Weeks
Test
Validation and testing with small and large groups
Evaluation
Model revision based on false positive classifications and user feedbackuser
Implementation
- Controlled trial on 40 UM Geography PPG students
- User Training
- Data collection

Figure.1. ADDIE Research Flow

The subjects of this study involved 40 students of the Teacher Professional Education Program (PPG) at Universitas Negeri Malang in the 2025 academic year. Sampling was conducted using Purposive Sampling, with respondents characterized as PPG Geography Education students enrolled in the Learning Evaluation course. To test the system's effectiveness, respondents were divided into two groups in a quasi-experimental design: 20 students in the experimental group using Google Sheets integrated with Deep Learning, and 20 students in the control group using the conventional Excel method.

Qualitative and quantitative data were obtained directly from graduate students in the PPG program at Universitas Negeri Malang through three structured instruments (Ardiansyah et al., 2023; Nasution, 2023). First, an expert validation questionnaire to evaluate content validity and practicality. Second, a user response questionnaire trial for education practitioners and students using a Likert scale to assess Content Quality, Interactive Design and Affordability, Implementation in Learning, and Practical Needs. Third, implementation with students to measure the improvement in their skills in formulating questions based on Higher Order Thinking Skills (HOTS) before the pre-test and after the post-test. The open instrument produced qualitative data in the form of criticism and suggestions, while the closed instrument produced validity scores and learning outcomes, which were then analyzed for content validity using Aiken's V Formula:

$$V = \frac{\sum\left(\frac{r_i - r_{min}}{r_{max} - r_{min}}\right)}{n}$$

Explanation:
V : Validity coefficient
r_i : Score given by validator i
r_{min} : Lowest score on the rating scale
r_{max} : Highest score on the rating scale
n : Number of validators

Validity criteria V ≥ 0.75 (Ligasan & Doysabas, 2024), (b) quantitative statistical evaluation including Cronbach's Alpha reliability test ($\alpha \geq 0.7$) and paired sample t-test ($p < 0.05$), and (c) thematic analysis (Bujang et al., 2018) for qualitative data, facilitated by SPSS 26.0 software. The scores obtained from respondents' evaluations are quantitative data that will be converted to qualitative data by referring to the assessment range in Table 1 as follows.

Table.1. Content Criteria and Suitability

No	Description	Class
1	0%-20%	Very Unsuitable (STS)
2	21%-40%	Less Suitable (KL)
3	41%-60%	Undecided (R)
4	61%-80%	Suitable (L)
5	81%-100%	Very Good (SL)

Source: (Sumartini et al., 2020)

This study used the pretest and posttest scores of graduate students in the Geography PPG program at Malang State University for data analysis. Before conducting a paired-samples t-test, normality and homogeneity tests were performed to ensure the data met the necessary statistical assumptions. The pretest and posttest results were then evaluated using a paired sample t-test to test the research hypothesis. As a result, the t-test results were compared with the predetermined significance threshold.

4. Results and Discussion

4.1. Research Results

4.1.1. Analysis

Interviews with geography PPG lecturers revealed that learning evaluation courses still use conventional pedagogical approaches centered on instructors, resulting in decreased engagement and enthusiasm among geography students. Significant findings show that students have difficulty understanding the material in the learning outcome evaluation module and show a lack of interest in static media such as PowerPoint. This study provides a unique teaching tool, namely an evaluation tool that integrates deep learning-based Google Sheets to overcome these difficulties. This solution aims to increase learning enthusiasm and reduce boredom, while also serving as a means to assess the analytical abilities of PPG Geography students in formulating Higher Order

Thinking Skills questions objectively, thereby promoting a comprehensive understanding of evaluation concepts.

A user needs analysis was conducted to identify the evaluation tools needed by graduate students in the geography program (PPG) at Universitas Negeri Malang. A Google Form questionnaire was distributed to geography students at Malang State University to facilitate data collection. Based on the results of the needs analysis in Figure 2, the need to develop evaluation tools was confirmed by the significant gap between competency requirements and the availability of learning facilities. It was noted that 60% of students stated that conventional learning resources were no longer adequate to guide instrument development, and 50% of respondents experienced significant difficulties in constructing HOTS questions independently. Students felt that the old methods reduced their learning motivation by 70%. As a solution, the majority of respondents (92%) specifically wanted the integration of Google Sheets technology based on Deep Learning.

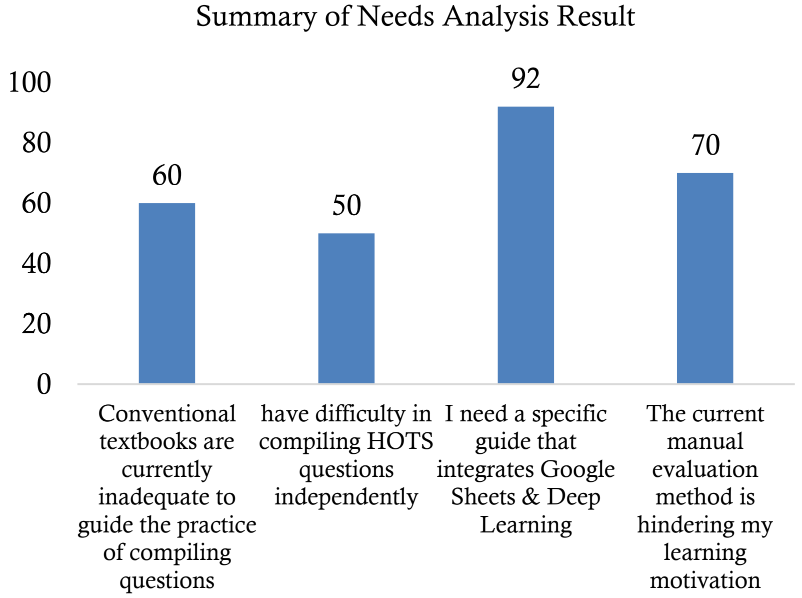


Figure.2. Results of Needs Analysis

4.1.2. Design

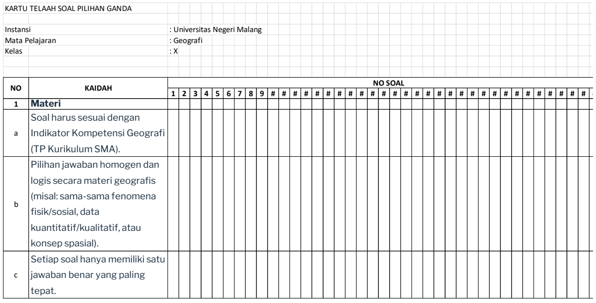
A Deep Learning-based Google Sheets evaluation tool was created to improve the learning experience of PPG students in formulating geography questions oriented towards Higher Order Thinking Skills. This process began with selecting an optimized platform as an effective educational tool. This involved designing a storyboard as a framework for the system interface, from the home page to the evaluation module, as well as developing content through the identification of core competencies, determination of material types, and selection of reference learning sources. The result is an integrated Deep Learning evaluation tool aligned with the research objectives that assesses the analytical skills of PPG students through a combination of Google Sheets and Deep Learning.

4.1.3. Development

The development stage of the Deep Learning-based Google Sheets evaluation tool focuses on creating and testing the feasibility of an interactive evaluation system. The process begins with formulating learning objectives and preparing materials in accordance with the syllabus. Next, the evaluation tool is developed based on the initial design, complete with a user manual for students.

The evaluation tool has several features, including review cards, introduction, cover, data entry, results analysis, deep question item analysis, remedial enrichment, remedial attendance list, and STEM (Science, Technology, Engineering, and Mathematics) deep-based remedial material. The final result is a structured evaluation tool based on competencies, complete with a questionnaire sheet to measure the validity of the user response tool.

The feature details are illustrated in Figure 3, which shows the customized review card sheet for the high school geography review card that has been created. Figure 4 presents the difficulty-level analysis, which aims to determine whether the question is easy, medium, or difficult. Figure 5 shows the cover page. Create the cover page according to your creativity. Figure 6 shows the analysis of questions tested in previous exams or other assignments. This is the result of an analysis of multiple-choice questions based on deep learning. It displays the identity of the question analysis. Fill in the form as instructed, and also fill in the answer key in the column marked with a red circle to make it structured. Figure 7 shows the results of the question analysis. Green indicates the top group with high scores, while red indicates the bottom group with low scores. Figure 8 shows the results of the question difficulty analysis. We can see which questions are difficult and which are not, and whether they have been revised or rejected/accepted. Figure 9 shows the remedial and enrichment display, where you can see the names of students undergoing remediation along with the suggestions they must follow according to the predetermined rules. Figure 10 shows the remedial attendance list, where students decide whether to do remedial work together or individually. However, they must still pay attention to the teacher's schedule, especially for geography lessons. Figure 11, the last page, is the remedial form. This form can be adjusted to each student's remedial needs. This form has been integrated with Google Gemini Artificial Intelligence for remedial recommendations based on Deep Learning material.

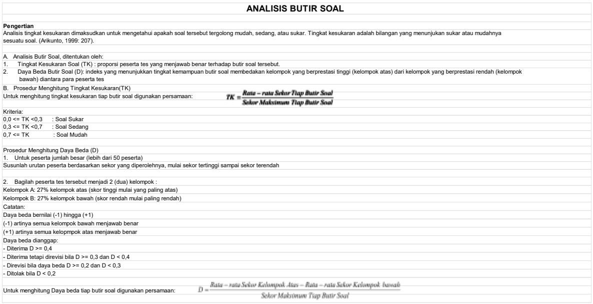


KARTU TELAAH SOAL PILIHAN GANDA

Instansi : Universitas Negeri Malang
Mata Pelajaran : Geografi
Kelas : X

NO	KAIDAH	NO SOAL																																			
		1	2	3	4	5	6	7	8	9	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#
1	**Materi**																																					
a	Soal harus sesuai dengan Indikator Kompetensi Geografi (TP Kurikulum SMA).																																					
b	Pilihan jawaban homogen dan logis secara materi geografis (misal: sama-sama fenomena fisik/sosial, data kuantitatif/kualitatif, atau konsep spasial).																																					
c	Setiap soal hanya memiliki satu jawaban benar yang paling tepat.																																					

Figure.3. Initial Display of the Review Card



ANALISIS BUTIR SOAL

Pengertian
Analisis tingkat kesukaran dimaksudkan untuk mengetahui apakah soal tersebut tergolong mudah, sedang, atau sukar. Tingkat kesukaran adalah bilangan yang menunjukan sukar atau mudahnya sesuatu soal. (Arikunto, 1999: 207).

A. Analisis Butir Soal, ditentukan oleh:
1. Tingkat Kesukaran Soal (TK) : proporsi peserta tes yang menjawab benar terhadap butir soal tersebut.
2. Daya Beda Butir Soal (D): indeks yang menunjukkan tingkat kemampuan butir soal membedakan kelompok yang berprestasi tinggi (kelompok atas) dari kelompok yang berprestasi rendah (kelompok bawah) diantara para peserta tes

B. Prosedur Menghitung Tingkat Kesukaran(TK)
Untuk menghitung tingkat kesukaran tiap butir soal digunakan persamaan:

$$TK = \frac{\text{Rata – rata Sekor Tiap Butir Soal}}{\text{Sekor Maksimum Tiap Butir Soal}}$$

Kriteria:
0,0 <= TK <0,3 : Soal Sukar
0,3 <= TK <0,7 : Soal Sedang
0,7 <= TK : Soal Mudah

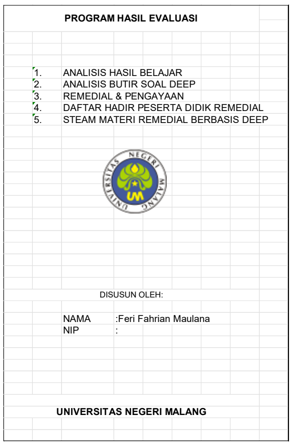
Prosedur Menghitung Daya Beda (D)
1. Untuk peserta jumlah besar (lebih dari 50 peserta)
Susunlah urutan peserta berdasarkan sekor yang diperolehnya, mulai sekor tertinggi sampai sekor terendah

2. Bagilah peserta tes tersebut menjadi 2 (dua) kelompok :
Kelompok A: 27% kelompok atas (skor tinggi mulai yang paling atas)
Kelompok B: 27% kelompok bawah (skor rendah mulai paling rendah)
Catatan:
Daya beda bernilai (-1) hingga (+1)
(-1) artinya semua kelompok bawah menjawab benar
(+1) artinya semua kelopmpok atas menjawab benar
Daya beda dianggap:
- Diterima D >= 0,4
- Diterima tetapi direvisi bila D >= 0,3 dan D < 0,4
- Direvisi bila daya beda D >= 0,2 dan D < 0,3
- Ditolak bila D < 0,2

Untuk menghitung Daya beda tiap butir soal digunakan persamaan:

$$D = \frac{\text{Rata – rata Sekor Kelompok Atas – Rata – rata Sekor Kelompok bawah}}{\text{Sekor Maksimum Tiap Butir Soal}}$$

Figure.4. Introduction to Item Analysis



PROGRAM HASIL EVALUASI

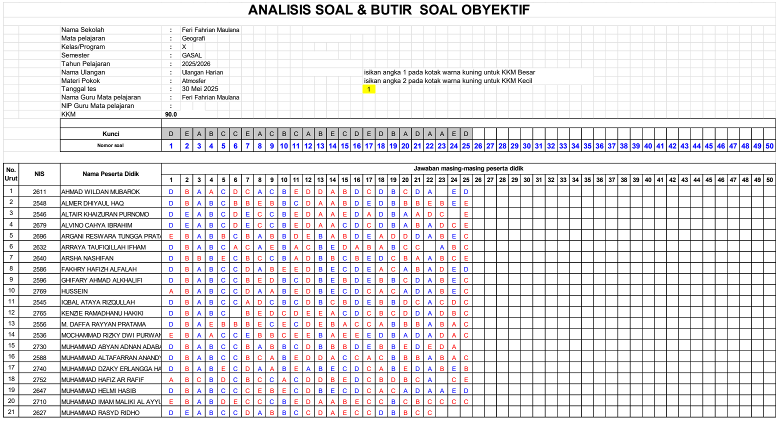
1. ANALISIS HASIL BELAJAR
2. ANALISIS BUTIR SOAL DEEP
3. REMEDIAL & PENGAYAAN
4. DAFTAR HADIR PESERTA DIDIK REMEDIAL
5. STEAM MATERI REMEDIAL BERBASIS DEEP

DISUSUN OLEH:

NAMA :Feri Fahrian Maulana
NIP :

UNIVERSITAS NEGERI MALANG

Figure.5. Cover Display



ANALISIS SOAL & BUTIR SOAL OBYEKTIF

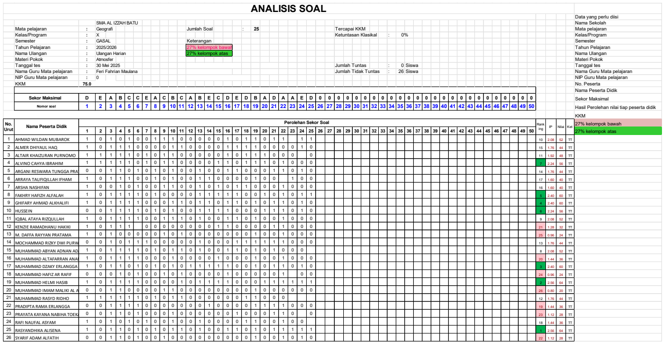
Nama Sekolah	:	Feri Fahrian Maulana
Mata pelajaran	:	Geografi
Kelas/Program	:	X
Semester	:	GASAL
Tahun Pelajaran	:	2025/2026
Nama Ulangan	:	Ulangan Harian
Materi Pokok	:	Atmosfer
Tanggal tes	:	30 Mei 2025
Nama Guru Mata pelajaran	:	Feri Fahrian Maulana
NIP Guru Mata pelajaran	:	
KKM	90.0	

isikan angka 1 pada kotak warna kuning untuk KKM Besar
isikan angka 2 pada kotak warna kuning untuk KKM Kecil
1

Kunci	D	E	A	B	C	C	E	A	C	B	C	A	B	E	C	D	E	D	B	A	D	A	A	E	D																									
Nomor soal	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50

No. Urut	NIS	Nama Peserta Didik	Jawaban masing-masing peserta didik																																																	
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	2611	AHMAD WILDAN MUBAROK	D	B	A	A	C	D	C	A	C	B	E	D	D	A	B	D	C	D	B	C	D	A		E	D																									
2	2548	ALMER DHIYAUL HAQ	D	B	A	B	C	B	B	E	B	B	C	D	A	A	B	D	E	D	B	B	B	E	B	E	E																									
3	2546	ALTAIR KHAIZURAN PURNOMO	D	E	A	B	C	D	E	C	C	B	E	D	A	A	E	D	A	D	B	A	A	D	C		E																									
4	2679	ALVINO CAHYA IBRAHIM	D	E	A	B	C	D	E	C	C	B	E	D	A	A	C	D	C	D	B	A	B	A	D	C	E																									
5	2696	ARGANI RESWARA TUNGGA PRAT/	E	B	A	B	B	C	B	A	B	B	D	E	B	A	B	D	E	A	D	D	D	A	B	E	C																									
6	2632	ARRAYA TAUFIQILLAH IFHAM	D	B	A	B	C	A	C	A	E	B	A	C	B	E	D	A	B	A	B	C	C		A	B	C																									
7	2640	ARSHA NASHIFAN	D	B	B	B	E	C	B	C	C	B	A	D	B	B	C	B	E	D	C	B	A	A	B	C	E																									
8	2586	FAKHRY HAFIZH ALFALAH	D	B	A	B	C	C	D	A	B	E	E	D	B	E	C	D	E	A	C	A	B	A	D	E	D																									
9	2596	GHIFARY AHMAD ALKHALIFI	D	B	A	B	C	C	B	E	D	B	C	D	B	E	B	D	E	B	B	C	D	A	B	E	C																									
10	2769	HUSSEIN	A	B	A	B	C	C	D	A	A	B	E	D	B	E	C	D	C	A	C	A	D	A	B	E	C																									
11	2545	IQBAL ATAYA RIZQULLAH	D	B	A	B	C	C	A	D	C	B	C	D	B	C	B	D	E	B	B	D	C	A	C	D	C																									
12	2765	KENZIE RAMADHANU HAKIKI	D	B	A	B	C		B	E	D	C	D	E	E	A	C	D	C	B	C	D	D	A	D	B	C																									
13	2556	M. DAFFA RAYYAN PRATAMA	D	B	A	E	B	B	B	E	C	E	C	D	E	B	A	C	C	A	B	B	B	A	B	A	C																									
14	2536	MOCHAMMAD RIZKY DWI PURWAN	E	B	A	A	C	C	E	B	B	C	E	E	B	A	E	E	E	D	B	A	D	A	D	A	C																									
15	2730	MUHAMMAD ABYAN ADNAN ADAB/	D	B	A	B	C	C	B	A	B	B	C	D	B	B	B	D	E	B	B	E	D	E	D	A																										
16	2588	MUHAMMAD ALTAFARRAN ANANDY	D	B	A	B	C	C	B	C	A	B	E	D	D	A	C	C	A	C	B	B	B	A	B	A	C																									
17	2740	MUHAMMAD DZAKY ERLANGGA H/	D	B	A	B	E	C	D	A	A	B	E	A	B	E	C	D	C	A	B	E	D	A	B	E	B																									
18	2752	MUHAMMAD HAFIZ AR RAFIF	A	B	C	B	D	C	B	C	C	A	C	D	D	B	E	D	C	B	D	B	C	A		C	E																									
19	2647	MUHAMMAD HELMI HASIB	D	B	A	B	C	C	C	E	B	E	C	D	B	E	C	D	C	A	C	A	D	A	A	E	D																									
20	2710	MUHAMMAD IMAM MALIKI AL AYYU	E	B	A	B	D	E	C	C	C	B	E	D	A	A	B	E	C	C	B	C	B	C	C	C	C																									
21	2627	MUHAMMAD RASYD RIDHO	D	E	A	B	C	C	D	A	B	B	C	C	D	A	E	C	C	D	B	B	C	C																										

Figure.6. Analysis of Questions and Objective Items



ANALISIS SOAL

SMA AL IZZAH BATU
Mata pelajaran : Geografi
Kelas/Program : X
Semester : GASAL
Tahun Pelajaran : 2025/2026
Nama Ulangan : Ulangan Harian
Materi Pokok : Atmosfer
Tanggal tes : 30 Mei 2025
Nama Guru Mata pelajaran : Feri Fahrian Maulana
NIP Guru Mata pelajaran : 0
KKM 75.0

Jumlah Soal : 25
Keterangan
🟥 27% kelompok bawah
🟩 27% kelompok atas

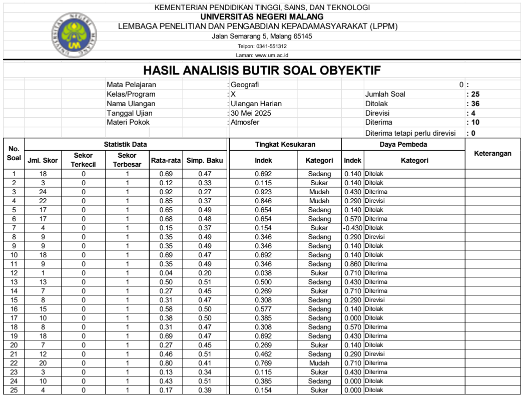
Tercapai KKM
Ketuntasan Klasikal : 0%
Jumlah Tuntas : 0 Siswa
Jumlah Tidak Tuntas : 26 Siswa

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Sekor Maksimal	D	E	A	B	C	C	E	A	C	B	C	A	B	E	C	D	E	D	B	A	D	A	A	E	D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nomor soal	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50

No. Urut	Nama Peserta Didik	Perolehan Sekor Soal																																																		Rank ing	IP	Nilai	Ket
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50				
1	AHMAD WILDAN MUBAROK	1	0	1	0	1	0	0	1	1	1	0	0	0	0	0	1	0	1	1	0	1	1		1	1																										10	2.08	52	TT
2	ALMER DHIYAUL HAQ	1	0	1	1	1	0	0	0	0	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	0																										15	1.76	44	TT
3	ALTAIR KHAIZURAN PURNOMO	1	1	1	1	1	0	1	0	1	1	0	0	0	0	0	1	0	1	1	1	0	0	0		0																										11	1.92	48	TT
4	ALVINO CAHYA IBRAHIM	1	1	1	1	1	0	1	0	1	1	0	0	0	0	1	1	0	1	1	1	0	1	0	0	0																										7	2.24	56	TT
5	ARGANI RESWARA TUNGGA PRA	0	0	1	1	0	1	0	1	0	1	0	0	1	0	0	1	1	0	0	0	1	1	0	1	0																										14	1.76	44	TT
6	ARRAYA TAUFIQILLAH IFHAM	1	0	1	1	1	0	0	1	0	1	0	0	1	1	0	0	0	0	1	0	0		1	0	0																										17	1.60	40	TT
7	ARSHA NASHIFAN	1	0	0	1	0	1	0	0	1	1	0	0	1	0	1	0	1	1	0	0	0	1	0	0	0																										16	1.60	40	TT
8	FAKHRY HAFIZH ALFALAH	1	0	1	1	1	1	0	1	0	0	0	0	1	1	1	1	1	0	0	1	0	1	0	1	1																										5	2.40	60	TT
9	GHIFARY AHMAD ALKHALIFI	1	0	1	1	1	1	0	0	0	1	1	0	1	1	0	1	1	0	1	0	1	1	0	1	0																										4	2.40	60	TT
10	HUSSEIN	0	0	1	1	1	1	0	1	0	1	0	0	1	1	1	1	0	0	0	1	1	1	0	1	0																										6	2.24	56	TT
11	IQBAL ATAYA RIZQULLAH	1	0	1	1	1	1	0	0	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0																										9	2.08	52	TT
12	KENZIE RAMADHANU HAKIKI	1	0	1	1	1		0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0																										21	1.28	32	TT
13	M. DAFFA RAYYAN PRATAMA	1	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0																										25	0.96	24	TT
14	MOCHAMMAD RIZKY DWI PURW	0	0	1	0	1	1	1	0	0	0	0	0	1	0	0	0	1	1	1	1	1	1	0	0	0																										13	1.76	44	TT
15	MUHAMMAD ABYAN ADNAN AD	1	0	1	1	1	1	0	1	0	1	1	0	1	0	0	1	1	0	1	0	1	0	0	0																											8	2.08	52	TT
16	MUHAMMAD ALTAFARRAN ANA	1	0	1	1	1	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0																										20	1.44	36	TT
17	MUHAMMAD DZAKY ERLANGGA	1	0	1	1	0	1	0	1	0	1	0	1	1	1	1	1	0	0	1	0	1	1	0	1	0																										3	2.40	60	TT
18	MUHAMMAD HAFIZ AR RAFIF	0	0	0	1	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	1		0	0																										24	0.96	24	TT
19	MUHAMMAD HELMI HASIB	1	0	1	1	1	1	0	0	0	0	1	0	1	1	1	1	0	0	0	1	1	1	1	1	1																										2	2.56	64	TT
20	MUHAMMAD IMAM MALIKI AL A	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0																										26	0.80	20	TT
21	MUHAMMAD RASYD RIDHO	1	1	1	1	1	1	0	1	0	1	1	0	0	0	0	0	0	1	1	0	0	0																													12	1.76	44	TT
22	PRADIPTA RAMA ERLANGGA	0	0	1	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0	0	0																										19	1.44	36	TT
23	PRAYATA KAYANA NABIHA TOEK	0	0	1	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	1	0		0																										23	1.12	28	TT
24	RAFI NAUFAL ASYAM	1	0	1	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	0	0	1	0	0																											18	1.44	36	TT
25	RASYANDHIKA ALISENA	1	0	1	1	0	1	0	1	0	1	1	0	1	0	0	1	1	0	1	0	1	1	1	1	1																										1	2.56	64	TT
26	SYARIF ADAM ALFATIH	0	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0																										22	1.12	28	TT

Data yang perlu diisi
Nama Sekolah
Mata pelajaran
Kelas/Program
Semester
Tahun Pelajaran
Nama Ulangan
Materi Pokok
Tanggal tes
Nama Guru Mata pelajaran
NIP Guru Mata pelajaran
No. Peserta
Nama Peserta Didik
Sekor Maksimal
Hasil Perolehan nilai tiap peserta didik
KKM
🟥 27% kelompok bawah
🟩 27% kelompok atas

Figure.7. Item Analysis in One Class



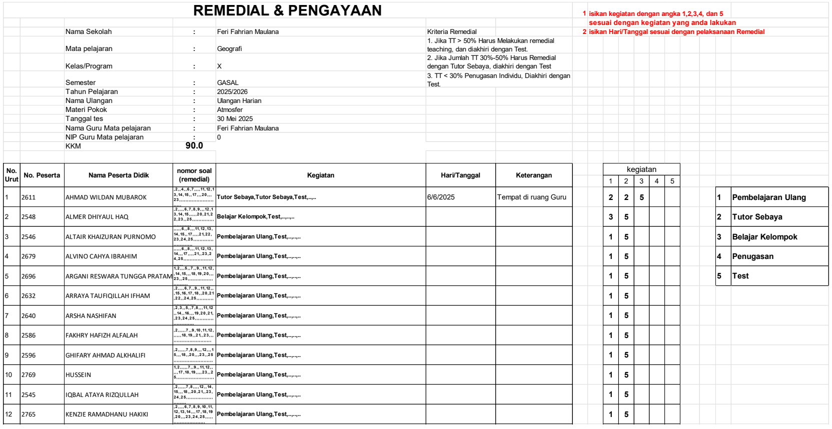
KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS NEGERI MALANG
LEMBAGA PENELITIAN DAN PENGABDIAN KEPADAMASYARAKAT (LPPM)
Jalan Semarang 5, Malang 65145
Telpon: 0341-551312
Laman: www.um.ac.id

HASIL ANALISIS BUTIR SOAL OBYEKTIF

Mata Pelajaran	: Geografi	0	:
Kelas/Program	: X	Jumlah Soal	: 25
Nama Ulangan	: Ulangan Harian	Ditolak	: 36
Tanggal Ujian	: 30 Mei 2025	Direvisi	: 4
Materi Pokok	: Atmosfer	Diterima	: 10
		Diterima tetapi perlu direvisi	: 0

No. Soal	Statistik Data					Tingkat Kesukaran		Daya Pembeda		Keterangan
	Jml. Skor	Sekor Terkecil	Sekor Terbesar	Rata-rata	Simp. Baku	Indek	Kategori	Indek	Kategori	
1	18	0	1	0.69	0.47	0.692	Sedang	0.140	Ditolak	
2	3	0	1	0.12	0.33	0.115	Sukar	0.140	Ditolak	
3	24	0	1	0.92	0.27	0.923	Mudah	0.430	Diterima	
4	22	0	1	0.85	0.37	0.846	Mudah	0.290	Direvisi	
5	17	0	1	0.65	0.49	0.654	Sedang	0.140	Ditolak	
6	17	0	1	0.68	0.48	0.654	Sedang	0.570	Diterima	
7	4	0	1	0.15	0.37	0.154	Sukar	-0.430	Ditolak	
8	9	0	1	0.35	0.49	0.346	Sedang	0.290	Direvisi	
9	9	0	1	0.35	0.49	0.346	Sedang	0.140	Ditolak	
10	18	0	1	0.69	0.47	0.692	Sedang	0.140	Ditolak	
11	9	0	1	0.35	0.49	0.346	Sedang	0.860	Diterima	
12	1	0	1	0.04	0.20	0.038	Sukar	0.710	Diterima	
13	13	0	1	0.50	0.51	0.500	Sedang	0.430	Diterima	
14	7	0	1	0.27	0.45	0.269	Sukar	0.710	Diterima	
15	8	0	1	0.31	0.47	0.308	Sedang	0.290	Direvisi	
16	15	0	1	0.58	0.50	0.577	Sedang	0.140	Ditolak	
17	10	0	1	0.38	0.50	0.385	Sedang	0.000	Ditolak	
18	8	0	1	0.31	0.47	0.308	Sedang	0.570	Diterima	
19	18	0	1	0.69	0.47	0.692	Sedang	0.430	Diterima	
20	7	0	1	0.27	0.45	0.269	Sukar	0.140	Ditolak	
21	12	0	1	0.46	0.51	0.462	Sedang	0.290	Direvisi	
22	20	0	1	0.80	0.41	0.769	Mudah	0.710	Diterima	
23	3	0	1	0.13	0.34	0.115	Sukar	0.430	Diterima	
24	10	0	1	0.43	0.51	0.385	Sedang	0.000	Ditolak	
25	4	0	1	0.17	0.39	0.154	Sukar	0.000	Ditolak	

Figure.8. Results of Objective Item Analysis



REMEDIAL & PENGAYAAN

1 isikan kegiatan dengan angka 1,2,3,4, dan 5
sesuai dengan kegiatan yang anda lakukan
2 isikan Hari/Tanggal sesuai dengan pelaksanaan Remedial

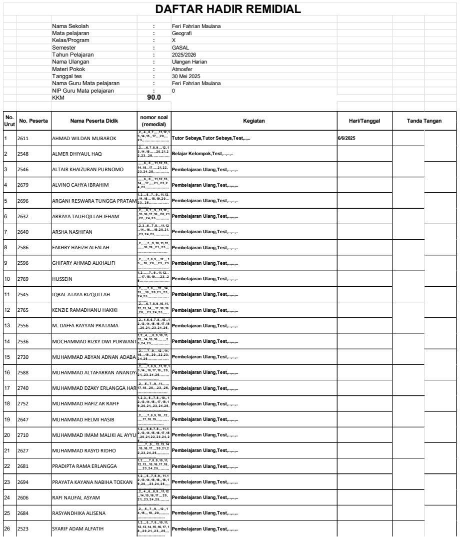
Nama Sekolah : Feri Fahrian Maulana
Mata pelajaran : Geografi
Kelas/Program : X
Semester : GASAL
Tahun Pelajaran : 2025/2026
Nama Ulangan : Ulangan Harian
Materi Pokok : Atmosfer
Tanggal tes : 30 Mei 2025
Nama Guru Mata pelajaran : Feri Fahrian Maulana
NIP Guru Mata pelajaran : 0
KKM 90.0

Kriteria Remedial
1. Jika TT > 50% Harus Melakukan remedial teaching, dan diakhiri dengan Test.
2. Jika Jumlah TT 30%-50% Harus Remedial dengan Tutor Sebaya, diakhiri dengan Test
3. TT < 30% Penugasan Individu, Diakhiri dengan Test.

No. Urut	No. Peserta	Nama Peserta Didik	nomor soal (remedial)	Kegiatan	Hari/Tanggal	Keterangan	kegiatan				
							1	2	3	4	5
1	2611	AHMAD WILDAN MUBAROK	,2,,4,,6,7,,,,11,12,13,14,15,,17,,,20,,,23,........	Tutor Sebaya,Tutor Sebaya,Test,..,..	6/6/2025	Tempat di ruang Guru	2	2	5		
2	2548	ALMER DHIYAUL HAQ	,2,,,,6,7,8,9,,,12,13,14,15,,,,,20,21,22,23,,25,........	Belajar Kelompok,Test,..,..,..			3	5			
3	2546	ALTAIR KHAIZURAN PURNOMO	,,,,,6,,8,,,11,12,13,14,15,,17,,,,21,22,23,24,25,........	Pembelajaran Ulang,Test,..,..,..			1	5			
4	2679	ALVINO CAHYA IBRAHIM	,,,,,6,,8,,,11,12,13,14,,,17,,,,21,,23,24,25,........	Pembelajaran Ulang,Test,..,..,..			1	5			
5	2696	ARGANI RESWARA TUNGGA PRATAM	1,2,,,5,,7,,9,,11,12,,14,15,,,18,19,20,,,23,,25,........	Pembelajaran Ulang,Test,..,..,..			1	5			
6	2632	ARRAYA TAUFIQILLAH IFHAM	,2,,,,6,7,,9,,11,12,,,15,16,17,18,,20,21,22,,24,25,........	Pembelajaran Ulang,Test,..,..,..			1	5			
7	2640	ARSHA NASHIFAN	,2,3,,5,,7,8,,,11,12,,14,,16,,,19,20,21,,23,24,25,........	Pembelajaran Ulang,Test,..,..,..			1	5			
8	2586	FAKHRY HAFIZH ALFALAH	,2,,,,,7,,9,10,11,12,,,,,,18,19,,21,,23,,,........	Pembelajaran Ulang,Test,..,..,..			1	5			
9	2596	GHIFARY AHMAD ALKHALIFI	,2,,,,,7,8,9,,,12,,,15,,,18,,20,,,23,,25........	Pembelajaran Ulang,Test,..,..,..			1	5			
10	2769	HUSSEIN	1,2,,,,,7,,9,,11,12,,,,,17,18,19,,,,23,,25,........	Pembelajaran Ulang,Test,..,..,..			1	5			
11	2545	IQBAL ATAYA RIZQULLAH	,2,,,,,7,8,,,,12,,14,15,,,18,,20,21,,23,24,25,........	Pembelajaran Ulang,Test,..,..,..			1	5			
12	2765	KENZIE RAMADHANU HAKIKI	,2,,,,6,7,8,9,10,11,12,13,14,,,17,18,19,20,,,23,24,25,........	Pembelajaran Ulang,Test,..,..,..			1	5			

1	Pembelajaran Ulang
2	Tutor Sebaya
3	Belajar Kelompok
4	Penugasan
5	Test

Figure.9. Remedial & Enrichment

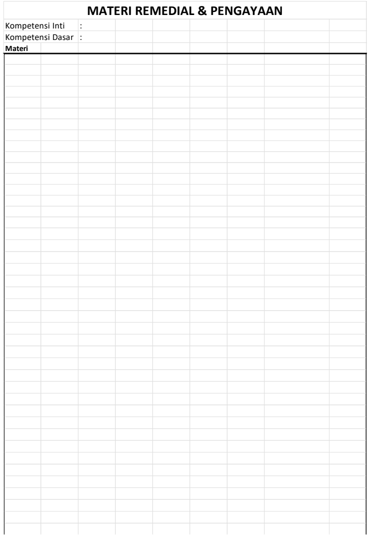


DAFTAR HADIR REMIDIAL

Nama Sekolah : Feri Fahrian Maulana
Mata pelajaran : Geografi
Kelas/Program : X
Semester : GASAL
Tahun Pelajaran : 2025/2026
Nama Ulangan : Ulangan Harian
Materi Pokok : Atmosfer
Tanggal tes : 30 Mei 2025
Nama Guru Mata pelajaran : Feri Fahrian Maulana
NIP Guru Mata pelajaran : 0
KKM 90.0

No. Urut	No. Peserta	Nama Peserta Didik	nomor soal (remedial)	Kegiatan	Hari/Tanggal	Tanda Tangan
1	2611	AHMAD WILDAN MUBAROK	,2,,4,,6,7,,,,11,12,13,14,15,,17,,,20,,,23,........	Tutor Sebaya,Tutor Sebaya,Test,..,..	6/6/2025	
2	2548	ALMER DHIYAUL HAQ	,2,,,,6,7,8,9,,,12,13,14,15,,,,,20,21,22,23,,25,........	Belajar Kelompok,Test,..,..,..		
3	2546	ALTAIR KHAIZURAN PURNOMO	,,,,,6,,8,,,11,12,13,14,15,,17,,,,21,22,23,24,25,........	Pembelajaran Ulang,Test,..,..,..		
4	2679	ALVINO CAHYA IBRAHIM	,,,,,6,,8,,,11,12,13,14,,,17,,,,21,,23,24,25,........	Pembelajaran Ulang,Test,..,..,..		
5	2696	ARGANI RESWARA TUNGGA PRATAM	1,2,,,5,,7,,9,,11,12,,14,15,,,18,19,20,,,23,,25,........	Pembelajaran Ulang,Test,..,..,..		
6	2632	ARRAYA TAUFIQILLAH IFHAM	,2,,,,6,7,,9,,11,12,,,15,16,17,18,,20,21,22,,24,25,........	Pembelajaran Ulang,Test,..,..,..		
7	2640	ARSHA NASHIFAN	,2,3,,5,,7,8,,,11,12,,14,,16,,,19,20,21,,23,24,25,........	Pembelajaran Ulang,Test,..,..,..		
8	2586	FAKHRY HAFIZH ALFALAH	,2,,,,,7,,9,10,11,12,,,,,,18,19,,21,,23,,,........	Pembelajaran Ulang,Test,..,..,..		
9	2596	GHIFARY AHMAD ALKHALIFI	,2,,,,,7,8,9,,,12,,,15,,,18,,20,,,23,,25........	Pembelajaran Ulang,Test,..,..,..		
10	2769	HUSSEIN	1,2,,,,,7,,9,,11,12,,,,,17,18,19,,,,23,,25,........	Pembelajaran Ulang,Test,..,..,..		
11	2545	IQBAL ATAYA RIZQULLAH	,2,,,,,7,8,,,,12,,14,15,,,18,,20,21,,23,24,25,........	Pembelajaran Ulang,Test,..,..,..		
12	2765	KENZIE RAMADHANU HAKIKI	,2,,,,6,7,8,9,10,11,12,13,14,,,17,18,19,20,,,23,24,25,........	Pembelajaran Ulang,Test,..,..,..		
13	2556	M. DAFFA RAYYAN PRATAMA	,2,,4,5,6,7,8,,10,,12,13,14,15,16,17,18,,20,21,,23,24,25,........	Pembelajaran Ulang,Test,..,..,..		
14	2536	MOCHAMMAD RIZKY DWI PURWANT	1,2,,4,,,,8,9,10,11,12,,14,15,16,,,,,,,23,24,25,........	Pembelajaran Ulang,Test,..,..,..		
15	2730	MUHAMMAD ABYAN ADNAN ADABA	,2,,,,,7,,9,,,12,,14,15,,,18,,20,,22,23,24,25,........	Pembelajaran Ulang,Test,..,..,..		
16	2588	MUHAMMAD ALTAFARRAN ANANDY	,2,,,,,7,8,9,,11,12,13,14,,16,17,18,,20,21,,23,24,25,........	Pembelajaran Ulang,Test,..,..,..		
17	2740	MUHAMMAD DZAKY ERLANGGA HAR	,2,,,5,,7,,9,,11,,,,,,17,18,,20,,,23,,25,........	Pembelajaran Ulang,Test,..,..,..		
18	2752	MUHAMMAD HAFIZ AR RAFIF	1,2,3,,5,,7,8,,10,,12,13,14,15,,17,18,19,20,21,,23,24,25,........	Pembelajaran Ulang,Test,..,..,..		
19	2647	MUHAMMAD HELMI HASIB	,2,,,,,7,8,9,10,,12,,,,,17,18,19,,,,,,........	Pembelajaran Ulang,Test,..,..,..		
20	2710	MUHAMMAD IMAM MALIKI AL AYYU	1,2,,,5,6,7,8,,,11,12,13,14,15,16,17,18,,20,21,22,23,24,25........	Pembelajaran Ulang,Test,..,..,..		
21	2627	MUHAMMAD RASYD RIDHO	,,,,,,,7,,9,,,12,13,14,15,16,17,,,20,21,22,23,24,25,........	Pembelajaran Ulang,Test,..,..,..		
22	2681	PRADIPTA RAMA ERLANGGA	1,2,,,,,7,8,9,10,11,12,13,,15,16,17,18,,,,,23,24,25,........	Pembelajaran Ulang,Test,..,..,..		
23	2694	PRAYATA KAYANA NABIHA TOEKAN	1,2,,,5,,7,8,9,,11,12,13,14,15,16,,18,19,20,,,23,24,25,........	Pembelajaran Ulang,Test,..,..,..		
24	2606	RAFI NAUFAL ASYAM	,2,,4,,6,,8,9,,11,12,,14,15,16,17,,,20,21,,23,24,25,........	Pembelajaran Ulang,Test,..,..,..		
25	2684	RASYANDHIKA ALISENA	,2,,,5,,7,,9,,,12,,14,15,,,18,,20,........	Pembelajaran Ulang,Test,..,..,..		
26	2523	SYARIF ADAM ALFATIH	1,2,,,5,,7,8,,10,11,12,13,14,15,16,17,18,,20,21,,23,,25,........	Pembelajaran Ulang,Test,..,..,..		

Figure.10. Remedial Attendance List



MATERI REMEDIAL & PENGAYAAN

Kompetensi Inti :
Kompetensi Dasar :
Materi

Figure.11. Deep AI Remedial Material

The content validation process involved two subject matter experts who assessed three main indicators, namely curriculum suitability, depth of concept, and reference currency. Based on Figure 12, the post-revision material aspect recorded an average percentage score of 90%. According to the Likert scale feasibility interpretation standard in Table 1, this score is categorized as 'Very Feasible' and does not require major revisions. This high validity indicates that the material has been explicitly, accurately, and relevantly compiled to meet the competency needs of pre-service geography teacher education students. This is crucial, as unstructured material can trigger excessive abstraction and hinder understanding retention (Setiawan et al., 2023). These findings empirically validate that the material is ready for limited-scale testing with practitioners and students, although some minor refinements are still recommended before full adoption.

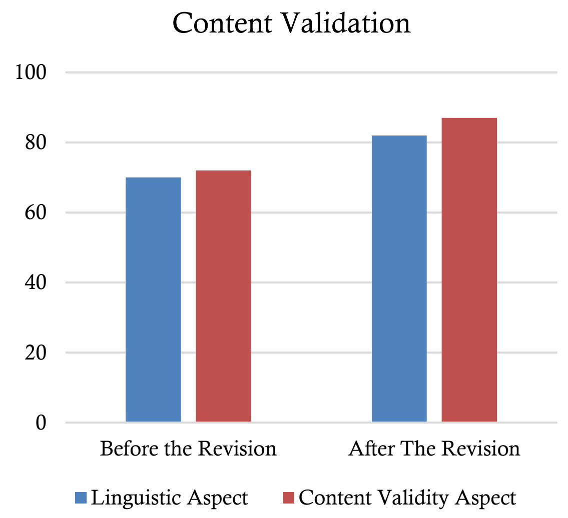


Figure.12. Material Validation

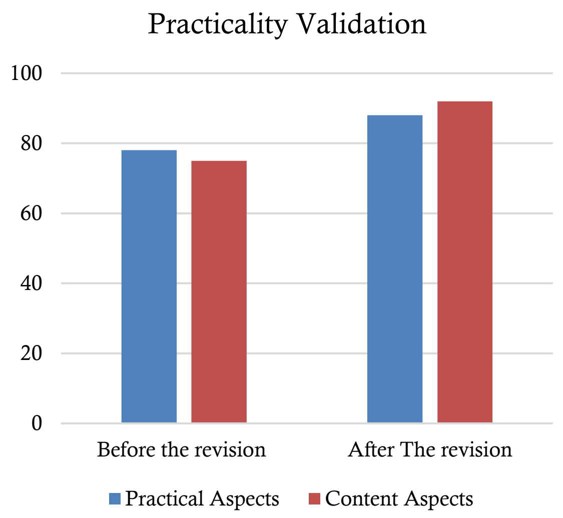


Figure.13. Practicality Validation

Based on the data visualization in Figure 13, the practicality test of this Deep Learning-based evaluation tool resulted in an average score of 88.5%. When converted according to the Likert scale guidelines in Table 1, this achievement falls into the Very Practical category. This high positive response indicates three main indicators, namely user-friendliness, AI analysis feature navigation flow, and system stability, which have been tested and declared very feasible to proceed to the field trial stage with a wider range of respondents. This technology provides a more immersive learning experience through visualization, helping students understand evaluation information through relevant images. This achievement is in line with the cone of experience theory (Dale, 1969), which emphasizes the importance of multisensory learning in improving teaching and learning methods.

The use of deep learning-based evaluation tools is expected to improve the experience of Geography PPG students in creating Higher Order Thinking Skills questions. Evaluation tools enable students to understand the process of question creation in depth (Sandika et al., 2025). This tool also facilitates students' understanding and development of skills in creating Higher Order Thinking Skills questions, enabling them to create deep, learning-based questions for high school students. This tool has been verified and is ready for testing with education practitioners and students of the Geography Education Master's Program (PPG).

4.1.4. Implementation

After undergoing expert validation during the development stage, a deep learning-based evaluation tool integrated with Google Sheets was applied to users to assess the analytical abilities of PPG students in compiling Higher Order Thinking Skills questions. The application involved lecturers as education practitioners and 20 postgraduate PPG students as the experimental group. The trial by practitioners focused on ease of use, the relevance of the model's analytical feedback, and educators' acceptance of the Google Sheet-based workflow. The assessment was conducted using a questionnaire containing 10 statements accompanied by comment and suggestion columns. The aspects assessed included language clarity, the suitability of system components such as Google Sheet templates, the mechanism for uploading Higher Order Thinking Skills items, the deep learning model inference results dashboard, the regularity of the material structure and assessment flow, and the overall quality of presentation and reporting.

This stage is intended to assess student responses to a deep learning-based Google Sheet-integrated evaluation in PPG learning, specifically whether the system can improve students' understanding of learning evaluation and prepare Higher Order Thinking Skills questions for students. The tested material emphasizes interactive visual presentations that are relevant to students' needs, ranging from question-item input templates in Google Sheets to analytical dashboards of model-prediction results, thereby making the learning process more concise and focused. This implementation not only deepens students' conceptual understanding but also strengthens lecturers' roles as facilitators who guide the interpretation of automatic feedback, while fostering students' learning independence in accordance with constructivist principles.

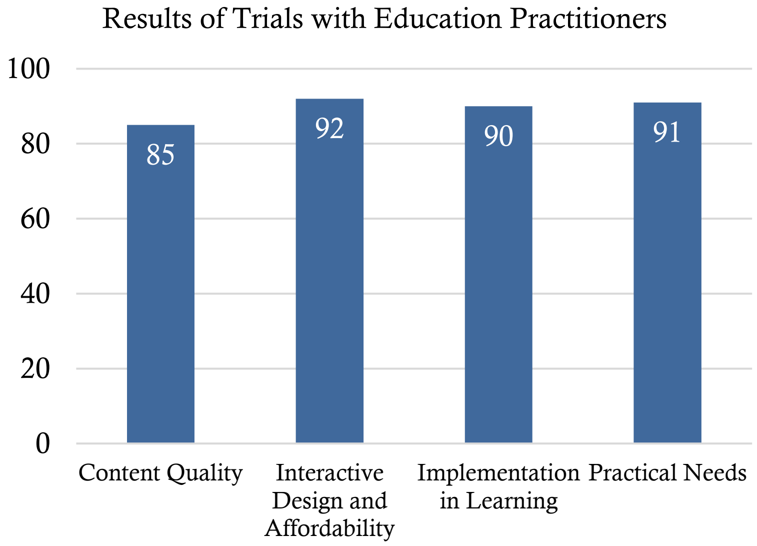


Figure.14. Graph of Test Results for Education Practitioners

Figure 14 shows the results of trials conducted by education practitioners, with an average feasibility of 95%, placing the deep learning-based Google Sheet-integrated evaluation tool in the highly feasible category. These findings confirm that the solution developed, ranging from the HOTS question input template in Google Sheet to the analytical dashboard of the model results, meets the quality criteria to proceed to a more extensive testing phase and is ready to be implemented for PPG students.

Data collection was carried out during offline learning in three sessions guided by PPG lecturers. Before the trial, students were directed to access the deep learning-based Google Sheet integrated evaluation tool on their respective devices. Next, the evaluation tool was used in accordance with the planned learning evaluation material flow. After all sessions were completed, students filled out a questionnaire containing statements to collect data related to the feasibility and effectiveness of the integrated system in supporting the measurement of their analytical skills in compiling Higher Order Thinking Skills questions.

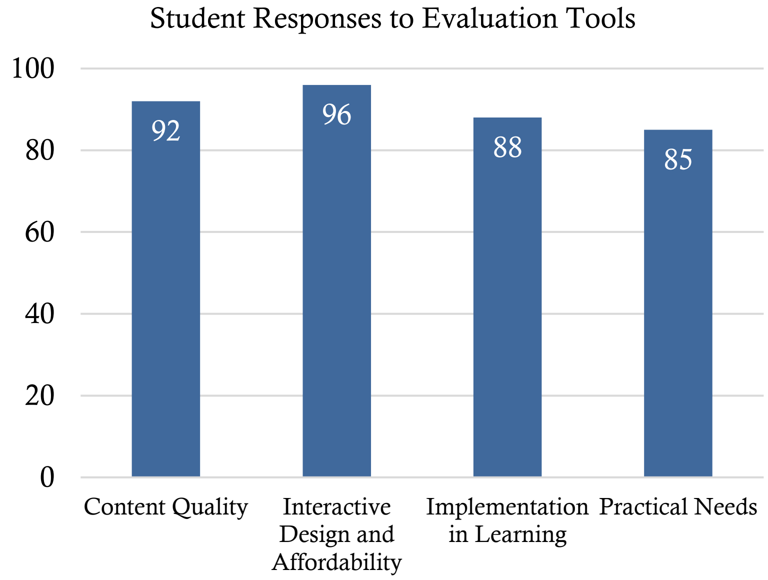


Figure.15. Graph of Student Responses to the Evaluation Tool

Referring to Figure 15, the deep learning-based Google Sheet integrated evaluation tool obtained an average feasibility of 90%, which places it in the highly feasible category at the trial use stage. Student responses indicate that the interface and workflow, from inputting Higher Order Thinking Skills items in Google Sheets to receiving analytical feedback from the model, are engaging, increase learning motivation, and facilitate understanding of the evaluation concept and how it works. This solution is considered appropriate for use in teaching evaluation material in the PPG program while also supporting the measurement of students' analytical skills in compiling HOTS questions.

4.1.5. Evaluation

Evaluation, as the final stage of the ADDIE model, is used to assess the achievement of the objectives of the evaluation tool that integrates deep learning-based Google Sheets into the Learning Evaluation material at PPG. The evaluation is carried out in layers at each phase, namely: First, Analysis, where researchers examine the results of mapping students' complexity and needs to ensure that the solution is relevant to the PPG context. The design is reviewed in terms of learning objective formulation, material development, questionnaire and test instrument design, and interaction flow storyboard from filling in Higher Order Thinking Skills items to result visualization, so that all elements support learning effectiveness. Development: expert validators assess the questionnaire's feasibility and identify shortcomings, such as typos, inaccurate illustrations, and the need to enrich the material to encourage higher-level thinking. Implementation stage: the questionnaire is filled out by practitioners and PPG students after using the evaluation tool. Practitioners gave positive responses regarding its effectiveness but suggested variations in application scenarios, while students reported an increase in understanding despite the guidebook's lack of detailed images that could be enlarged. This series of evaluations ensures that the development follows the ADDIE principles, resulting in effective teaching materials and evaluation mechanisms for Higher Order Thinking Skills that are tailored to student needs.

4.1.6. Effectiveness Analysis of Evaluation Tools

The effectiveness test was designed to assess the success of the deep learning-based Google Sheet integrated evaluation tool in improving the learning outcomes and analytical skills of PPG students in developing Higher Order Thinking Skills questions. The design used was a pretest-posttest with experimental and control groups. In the first meeting, all students underwent a pretest to map their initial abilities. In the second meeting, the experimental group learned using a Google Sheet-based workflow, including entering Higher Order Thinking Skills items, automatic processing by a deep learning model, and analytical feedback, while the control group followed conventional learning using printed teaching materials. In the third meeting, both groups took a posttest to test the improvement in understanding and learning achievement after the intervention, comparing the impact of the deep learning-based Google Sheet evaluation tool approach with the conventional method.

Table.2. Normality Test

Treatment	Sig.
Experimental Pretest	0.095
Control Pretest	0.136
Experimental Posttest	0.490
Control Posttest	0.198

The learning outcome data on the deep learning-based Google Sheet evaluation tool meet the normality assumption (Table 2). Using the Kolmogorov-Smirnov test, the experimental group showed pretest and posttest significance values of 0.095 and 0.490, respectively, while the control group showed pretest and posttest significance values of 0.136 and 0.195, respectively. All Sig. values were > 0.05, so the distributions of pretest and posttest data in both groups were deemed normal and suitable for further inferential analysis.

Table.3. Homogeneity Test

	Levene Statistic	df1	df2	Sig.
Effectiveness of Evaluation Tools	0.021	1	68	0.845

The results of the homogeneity test on the pretest and posttest data in both the experimental and control groups showed a significance value of 0.845. Based on statistical testing criteria, because the significance value of 0.845 was> 0.05, the data variances in both groups were deemed homogeneous (Table 2). With the prerequisites of homogeneity and normality of the data previously tested, statistical analysis can proceed using a parametric test (t-test) to test the hypothesis of the effectiveness of the Deep Learning-based Google Sheet evaluation tool on PPG students' ability to compile HOTS questions.

The results of the paired t-test in SPSS show a two-tailed p-value of 0.045 (< 0.05), so H_0 is rejected, and Ha is accepted. These findings confirm that evaluation tools that integrate deep learning-based Google Sheets are effective in improving learning outcomes and in supporting PPG students' analytical skills in compiling HOTS questions, compared with before the intervention. This integrated solution is worthy of implementation in educational evaluation learning in the Teacher Professional Education Program

4.2. Discussion

The use of Google Sheets and deep learning integration in this study not only serves as a variation in learning media but also plays a major role as a precision measurement tool for analytical skills. Technically, this system works by diagnosing the sentence structure in the questions to automatically determine their cognitive level, whether HOTS or LOTS. Direct feedback from AI trains students' logic to distinguish, organize, and compile complex question elements. Thus, the improvement of student competence is based on clear and measurable cognitive parameters, not merely an increase in motivation. In line with this, digital media can encourage active participation, increase enthusiasm, and positively impact learning outcomes (Muzammil & Izzah, 2025). The implementation of interactive formats, such as evaluation tools, significantly increases motivation and academic achievement, making them a relevant reference for strengthening the design of deep learning-based Google Sheets for Learning Evaluation materials in PPG (Banda & Nzabahimana, 2022).

The development of digital evaluation tools in this study is manifested as a deep learning-based Google Sheet evaluation tool for compiling Higher Order Thinking Skills questions, emphasizing the crucial role of web-based learning design. Microlearning through web pages helps students reflect on the knowledge they have learned while consolidating it in structured micro-learning units (Lawson et al., 2024), an approach that is in line with the needs of Learning Evaluation material. This involves understanding concepts in depth and then applying them practically to test items. In line with this, presenting material with digital tools makes the learning process more enjoyable, reduces boredom, and facilitates the understanding of complex concepts, a finding that is relevant for designing input flows in Google Sheets, automatic feedback from deep learning models, and user-friendly analytical displays (Shone et al., 2018).

In line with previous findings, evaluation tools that integrate deep learning-based Google Sheets as an extension of interactive approaches are not only effective in reducing boredom but also proven to significantly increase student motivation and learning achievement. Its visual-interactive design encourages active student participation by entering Higher Order Thinking Skills items into Google Sheets, receiving analytical feedback from the model, and reflecting on improvements iteratively. This more engaging and participatory learning experience aligns with the mainstream of web-based learning, which is recognized as innovative and effective, making it relevant for application in Learning Evaluation materials in PPG to strengthen students' analytical skills in developing Higher Order Thinking Skills questions.

5. Conclution

Based on the findings of the research, the design of an evaluation tool that integrates Google Sheets based on deep learning is in line with the results of the analysis of complexity and user needs in the Learning Evaluation material at PPG. Statistical testing using a paired t-test showed that the Sig. value was $0.000 < 0.05$, so Ha was accepted, and H_0 was rejected. This confirms the effectiveness of deep learning-based Google Sheet evaluation tools for students in the experimental class, with significantly greater gains than in the control class, as reflected in the larger difference between pretest and post-test scores. In the field, several obstacles arose, such as dependence on Google Sheets and connectivity, which could potentially affect input quality and synchronization stability; limited coverage of items and sub-materials, so that generalization to other Higher Order Thinking Skills domains has not been tested; insufficient transparency of model decisions on the dashboard to explain the reasons for assessment; and minimal adaptive formative evaluation and collaboration features in Google Sheets. Future recommendations include expanding the material, adding a module explaining model decisions, enriching adaptive collaboration and evaluation features, strengthening data governance, and replicating across institutions and topics. This is expected to improve the accuracy, accountability, and effectiveness of this approach in measuring and strengthening the analytical abilities of PPG students in compiling Higher Order Thinking Skills questions.

6. References

Akdemir, E., & Barışçı, N. (2024). A review on deep learning applications with semantics. *Expert Systems with Applications*, *251*(1), 124029. https://doi.org/https://doi.org/10.1016/j.eswa.2024.124029

Aniek, O., Kusuma, S., Gede Ratnaya, I., & Candiasa, I. M. (2022). Evaluasi Pembelajaran Hybrid Pada Institusi Pendidikan Instiki Dengan Discrepancy Evaluation Model. *Lembaga Penjaminan Mutu*, *8*(2), 217-227. https://doi.org/10.25078/jpm.v8i02.1509

Ardiansyah, Risnita, & Jailani, M. S. (2023). Teknik Pengumpulan Data Dan Instrumen Penelitian Ilmiah Pendidikan Pada Pendekatan Kualitatif dan Kuantitatif. *Jurnal IHSAN : Jurnal Pendidikan Islam*, *1*(2), 1–9. https://doi.org/10.61104/ihsan.v1i2.5

Banda, H. J., & Nzabahimana, J. (2022). The Impact of Physics Education Technology (PhET) Interactive Simulation-Based Learning on Motivation and Academic Achievement Among Malawian Physics Students. *Journal of Science Education and Technology*, *32*, 127–141. https://doi.org/10.1007/s10956-022-10010-3

Bujang, M., Omar, E., & Baharum, N. (2018). A Review on Sample Size Determination for Cronbach's Alpha Test: A Simple Guide for Researchers. *The Malaysian Journal of Medical Sciences : MJMS*, *25*(6), 85–99. https://doi.org/10.21315/mjms2018.25.6.9

Chen, G., & Han, Z. (2025). The rise of deep learning: AI and engineering applications under the spotlight of the 2024 Nobel prize. *Intelligent Geoengineering*, *2*(1), 14–21. https://doi.org/10.1016/j.ige.2025.03.002

Dagal, I., Harrison, A., Ibrahim, A. W., & Mbasso, W. F. (2025). Comprehensive evaluation of data preprocessing and visualization techniques for enhanced classification and sampling. *Cluster Computing*, *28*(7), 476. https://doi.org/10.1007/s10586-025-05512-9

Dale, E. (1969). *Audiovisual methods in teaching*. New York: Holt, Rinehart and Winston Inc. The Dryden Press.

DeBell, T., Goertzen, L., Larson, L., Selbie, W., Selker, J., & Udell, C. (2019). OPEnS hub: Real-time data logging, connecting field sensors to google sheets. *Frontiers in Earth Science*, *7*, 1-6. https://doi.org/10.3389/feart.2019.00137

Dos Santos, L. R., Behr, A., & Duarte, G. R. (2025). Recognizing accounting as a STEM discipline through professional skills in accounting information systems☆. *Journal of Accounting Education*, *72*. https://doi.org/10.1016/j.jaccedu.2025.100970

Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment: 'Breakthrough? Or buncombe and ballyhoo?' *Journal of Computer Assisted Learning*, *37*(5), 1207–1216. https://doi.org/10.1111/jcal.12577

Gkintoni, E., Antonopoulou, H., Sortwell, A., & Halkiopoulos, C. (2025). Challenging Cognitive Load Theory: The Role of Educational Neuroscience and Artificial Intelligence in Redefining Learning Efficacy. In *Brain Sciences, 15*(2), Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/brainsci15020203

Huang, Z., Fu, X., & Zhao, J. (2025). Research on AIGC-Integrated Design Education for Sustainable Teaching: An Empirical Analysis Based on the TAM and TPACK Models. *Sustainability (Switzerland)*, *17*(12). https://doi.org/10.3390/su17125497

Kacmaz, G., & Dubé, A. K. (2025). Measuring teacher scaffolding in game-based learning: Emotional and responsibility scaffolds lead while resources and previous game experience shape practices. *Computers and Education*, *237*. https://doi.org/10.1016/j.compedu.2025.105393

Kasanah, M., & Pratama, P. A. (2024). Taksonomi Tujuan Pendidikan dan Evaluasi Hasil Belajar. *Pedagogik: Jurnal Pendidikan dan Riset*, *2*(2), 146–162. https://ejournal.edutechjaya.com/index.php/pedagogik/article/view/593

Khoiriyah, Z., & Shaleh, S. (2025). Solusi Alternatif Atas Problematika dalam Mengimplementasikan Penilaian Higher Order Thingking Skills (HOTS). *Al-Madrasah Jurnal Pendidikan Madrasah Ibtidaiyah*, *9*(2), 656. https://doi.org/10.35931/am.v9i2.4855

Kunicki, Z. J., Zambrotta, N. S., Tate, M. C., Surrusco, A. R., Risi, M. M., & Harlow, L. L. (2019). Keep Your Stats in the Cloud! Evaluating the Use of Google Sheets to Teach Quantitative Methods. *Journal of Statistics Education*, *27*(3), 188–197. https://doi.org/10.1080/10691898.2019.1665485

Lawson, A. A., Ramage, K., Gibson, M. E. S., King, C., Imbo-Nloga, C. C., & French, A. V. (2024). Pediatric and Adolescent Gynecology WebEd: A Brief Report of an Underutilized Online Learning Module. *Journal of Pediatric and Adolescent Gynecology*, *37*(3), 371–374. https://doi.org/https://doi.org/10.1016/j.jpag.2024.01.166

Li, X., & Zhang, Y. (2025). Security application of intrusion detection model based on deep learning in english online education. *Alexandria Engineering Journal*, *124*, 582–590. https://doi.org/10.1016/j.aej.2025.03.051

Ligasan, J. M., & Doysabas, J. C. (2024). Enhancing news writing skills of grade 4 students through graphic organizer strategy. *Davao Research Journal*, *15*(4), 83-94. https://doi.org/10.59120/drj.v15i4.275

Ma, Y., Sun, X., & Ma, A. (2025). On the assessment and reliability of political and ideological education in colleges using deep learning methods. *Alexandria Engineering Journal*, *119*, 511–517. https://doi.org/10.1016/j.aej.2025.01.114

Maulana, F. F., Deffinika, I., Jazilah, F., & Pokhrel, L. B. (2023). Virtual Learning Media Exploration 3D View of the Konto River Watershed to Support 21st Century Learning.

Jambura Geo Education Journal, *4*(2), 133–146. https://doi.org/10.34312/jgej.v4i2.21399

Maxnun, L., Kristiani, K., & Sulistyaningrum, C. D. (2024). Development of hots-based cognitive assessment instruments: ADDIE model. *Journal of Education and Learning (EduLearn)*, *18*(2), 489–498. https://doi.org/10.11591/edulearn.v18i2.21079

Mir, S., Shah, S. A., & Bhat, M. S. (2024). Evaluation of disaster preparedness in educational infrastructure of remote geographies using Analytical Hierarchy process. *International Journal of Disaster Risk Reduction*, *108*. https://doi.org/10.1016/j.ijdrr.2024.104514

Misnawati, M., Junari, J., Teibang, D., Ilham, I., & Luthfiyah, L. (2025). Evaluasi Hasil Asesmen Melalui Pemberian Umpan Balik dalam Tes Formatif sebagai Tolak Ukur Hasil Belajar Siswa. *JIIP-Jurnal Ilmiah Ilmu Pendidikan*, *8*(2), 2236–2242. https://doi.org/10.54371/jiip.v8i2.6836

Muzammil, M., & Izzah, S. A. (2025). Penerapan Pembelajaran Berdiferensiasi Dengan Media Digital Untuk Meningkatkan Hasil Belajar Siswa. *Edukasiana: Jurnal Inovasi Pendidikan*, *4*(3), 1237–1246. https://doi.org/10.56916/ejip.v4i3.1407

Nasbi, I. (2017). MANAJEMEN KURIKULUM: Sebuah Kajian Teoritis. *Idaarah: Jurnal Manajemen Pendidikan*, *1*(2), 318–330. https://doi.org/10.24252/idaarah.v1i2.4274

Nasser, M., Arshad, N. I., Ali, A., Alhussian, H., Saeed, F., Da'u, A., & Nafea, I. (2025). A systematic review of multimodal fake news detection on social media using deep learning models. In *Results in Engineering, 26*, Elsevier B.V. https://doi.org/10.1016/j.rineng.2025.104752

Nasution, A. fattah. (2023). *Metode Penelitian kualitatif*. Harva creative. https://harfacreative.com/

Parra, F., Jacobs, A., & Trevino, L. L. (2021). Shippy Express: Augmenting accounting education with Google Sheets. *Journal of Accounting Education*, *56*. https://doi.org/10.1016/j.jaccedu.2021.100740

Rebman, C. M., Booker, Q. E., Wimmer, H., Levkoff, S., Mcmurtrey, M., & Powell, L. M. (2023). An Industry Survey of Analytics Spreadsheet Tools Adoption: Microsoft Excel vs Google Sheets. In *Information Systems Education Journal (ISEDJ), 21*(5). https://isedj.org/;https://iscap.infohttps://isedj.org/;https://iscap.info

Sandika, N. A., Basit, A., Putra, R. J., Nufrizal, A., Putra, B. A., Gunawan, R. G., Ameliya, D., & Utami, R. J. (2025). *Deep Learning dalam Pendidikan Tinggi: Strategi, Metode, dan Implementasi*. Pradina Pustaka.

Saputra, D. A., & Nofrion. (2022). Pengembangan LMS Berbasis Google Sites untuk Mengembangkan HOTS Siswa pada Mata Pelajaran Geografi. *Jurnal Pendidikan Ilmu Pengetahuan Sosial Indonesia*, *7*(3), 206–220. https://journal.stkipsingkawang.ac.id/index.php/JurnalPIPSI/article/view/3147

Setiawan, Z., Pustikayasa, I. M., Jayanegara, I. N., Setiawan, I. N. A. F., Putra, I. N. A. S., Yasa, I. W. A. P., Asry, W., Arsana, I. N. A., Chaniago, G. G., Wibowo, S. E., & others.

(2023). *Pendidikan Multimedia : Konsep dan Aplikasi pada era revolusi industri 4.0 menuju society 5.0*. PT. Sonpedia Publishing Indonesia. https://books.google.co.id/books?id=7YjhEAAAQBAJ

Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *2*(1), 41–50. https://doi.org/10.1109/TETCI.2017.2772792

Sugrah, N. U. (2020). Implementasi teori belajar konstruktivisme dalam pembelajaran sains. *Humanika*, *19*(2), 121–138. https://doi.org/10.21831/hum.v19i2.29274

Sumartini, S., Harahap, K. S., & Sthevany, S. (2020). Kajian Pengendalian Mutu Produk Tuna Loin Precooked Frozen Menggunakan Metode Skala Likert Di Perusahaan Pembekuan Tuna. *Aurelia Journal*, *2*(1), 29. https://doi.org/10.15578/aj.v2i1.9392

Sumy, D. F., Drakes, O. O., McBride, S. K., & Jenkins, M. R. (2023). Social vulnerability and geographic access barriers to earthquake early warning education in museums and other free choice learning environments. *International Journal of Disaster Risk Reduction*, *97*. https://doi.org/10.1016/j.ijdrr.2023.104011

Surahman, E., Soepriyanto, Y., Wedi, A., & Ulfa, S. (2024). *Online Assessment: Panduan Guru Sekolah Dasar*. Academia Publication.

Vanhees, C., Nijlunsing, J., Muijs, D., Crato, N., Wils, M., Wiliam, D., Surma, T., & Kirschner, P. A. (2025). The role of knowledge-rich curricula in promoting deep thinking and complex skill acquisition. *Learning and Individual Differences*, *121*. https://doi.org/10.1016/j.lindif.2025.102729

Wahyudi, S., Arifin, T., Mutia, C., & Aji, H. (2025). Integrating Publuu technology into interactive teaching materials as. In *Jurnal Inovasi Teknologi Pendidikan, 21*(1). https://doi.org/10.21831/jitp.v12i1.78293

Yudistira, E. N., & Kom, S. (2024). *Deep Learning: Teori, Contoh Perhitungan, dan Implementasi*. Deepublish.