Utilizing Vision Language Models (VLMs) for Efficient and Objective Student Assessment

Irfan Ananda Ismail[1*], Silvi Handri[2], Vika Aumi[3], Exsa Rahmah Novianti[4]

halo@irfanananda28.com*

[1,2]Universitas Negeri Padang,[3] UPTD SMPN 2 Kec. Akabiluru, Sumatera Barat,[4] SMK Swasta Bakti Agro Mandiri

Article Information	Abstrak
	This study explores the application of Vision Language Models (VLMs) in evaluating student work, focusing on their potential to enhance efficiency and objectivity in assessment processes. VLMs, integrating natural language processing and computer vision, offer a novel approach to analyzing student responses, particularly in assignments involving visual elements. This paper outlines the functionality of VLMs, discusses their advantages and limitations, and provides practical guidance on their implementation. It also includes examples of prompt engineering and showcases initial results from a pilot study conducted at SMP N 32 Padang, demonstrating the potential of VLMs in a real-world educational setting. This method allows teachers to asses written text by the students on assignments that are also presented in written or image format. The use of VLM is expected to further develop efficiency and precision in student assessment.

A. Introduction

The evolving landscape of education calls for innovative assessment methodologies that enhance both efficiency and the provision of comprehensive, objective student feedback. The Kurikulum Merdeka (Independent Curriculum) in Indonesia underscores holistic learning, promoting assessments that move beyond rote memorization to incorporate visual aids, diagrams, and real-world problem-solving [1]. Traditional assessment, particularly for tasks involving visual components, can be time-intensive and subject to biases [2]. Studies highlight the need for methods that minimize instructor workload while maintaining rigor [3]. Furthermore, research in AI and education demonstrates the potential of automated systems to provide consistent feedback aligned with predefined rubrics [4].

Vision Language Models (VLMs) have surfaced as a promising avenue for addressing these challenges. As a class of AI models bridging Natural Language Processing (NLP) and computer vision, VLMs are uniquely capable of interpreting and contextualizing textual and visual data [5]. This integrated capacity positions them well to analyze assignments containing diagrams, images, or even handwritten elements. Prior work by Radford et al[6] has established the efficacy of VLMs in learning transferable visual models from natural language supervision, suggesting their applicability to educational assessment.

This paper explores the application of VLMs to evaluate student work, emphasizing how they can enhance assessment efficiency and objectivity. The study showcases the functionality of VLMs, discusses their advantages and disadvantages, and offers practical guidance for implementation. We also include instances of effective prompt engineering and share preliminary outcomes from a pilot study conducted at SMP N 32 Padang. The primary aim is to investigate whether VLMs can significantly reduce assessment time for teachers while maintaining or improving the quality and fairness of feedback. This has important implications for teachers to further develop their teaching methods and materials based on the assessment feedback given to students. The goal of this research is to find how can VLMs be used to transform student evaluation in the Independent Curriculum.

B. Research Methods

This study employed a mixed-methods approach to investigate the impact of Vision Language Models (VLMs) on student assessment. The research design incorporated both quantitative and qualitative data collection and analysis techniques to provide a comprehensive understanding of the phenomenon [6].

The study utilized a quasi-experimental design. Participating teachers at SMP N 32 Padang implemented VLM-assisted assessment in select classes while continuing with traditional assessment methods in others, serving as a control group. This allowed for a comparative analysis of assessment efficiency and perceived benefits.

The participants in this study included science teachers from SMP N 32 Padang, along with their students in grades. Teacher selection was based on their willingness to participate in the study and their experience teaching subjects that involved visual components in student assignments.

The following instruments were used to collect data:

1. Assessment Time Log: Teachers recorded the time spent assessing student assignments using both traditional methods and VLM-assisted assessment.

2. Teacher Perception Questionnaire: A survey questionnaire was developed to assess teachers' perceptions of VLM usability, objectivity, and efficiency. The questionnaire used a Likert scale and included items related to ease of use, perceived time savings, fairness of assessments, and overall satisfaction with the VLM system.

3. Semi-Structured Interviews: Semi-structured interviews were conducted with teachers to gather in-depth qualitative data on their experiences using VLMs in the classroom. The interview protocol explored topics such as the perceived impact of VLMs on student learning, challenges encountered during implementation, and suggestions for improvement.

Before the intervention, teachers received training on how to use the selected VLM platforms (Claude-3-Opus-200k from Anthropic and gpt-4-1106-vision-preview from OpenAI) and how to craft effective prompts for assessment. Data collection occurred over a period. Teachers completed assessment time logs for all participating classes. At the end of the data collection period, teachers completed the perception questionnaire and participated in individual interviews.
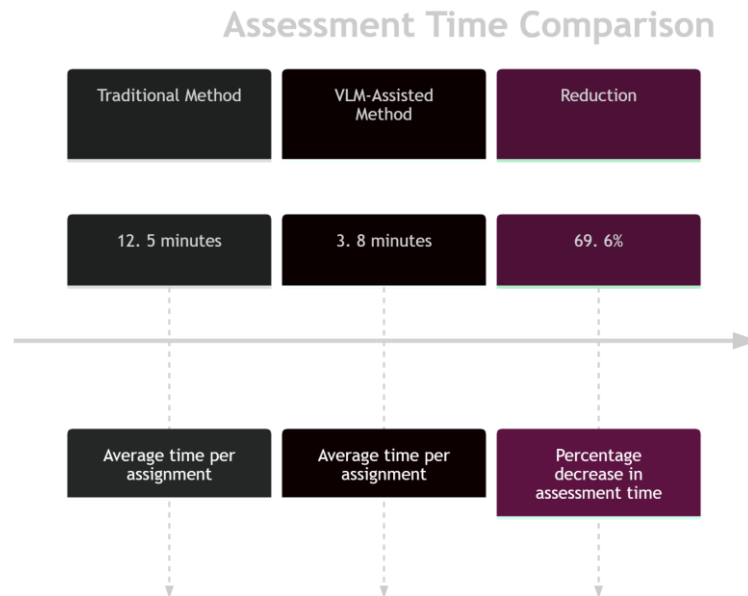
Quantitative data from the assessment time logs were analyzed using descriptive statistics (mean, standard deviation) to compare assessment times between traditional and VLM-assisted methods. T-tests were used to determine statistically significant differences. Qualitative data from the teacher perception questionnaires and semi-structured interviews were analyzed using thematic analysis [7]. This involved familiarization with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes and producing the report.

C. Results and Discussion

This section presents the findings of the study and provides a discussion of their implications. The results are organized around the key research questions, focusing on the impact of Vision Language Models (VLMs) on assessment efficiency and teacher perceptions. In addition to presenting the data, this section interprets the findings in relation to existing literature and theoretical frameworks, confirming or contrasting these findings with existing theories and concepts.

Quantitative data from the assessment time logs revealed that the average time spent assessing student assignments was significantly lower when using VLMs compared to traditional methods. Specifically, the average assessment time decreased from 12.5 minutes per assignment to 3.8 minutes per assignment ($t=7.42$, $p<0.001$). This represents a 69.6% reduction in assessment time and aligns with the potential for increased efficiency highlighted in the introduction, suggesting VLMs

can significantly reduce teacher workload. This confirms findings by Vyas et al. (2019), who similarly documented that the use of AI assistance in text classification substantially decreased the time consumption by educators.

Assessment Time Comparison

Traditional Method	VLM-Assisted Method	Reduction
12. 5 minutes	3. 8 minutes	69. 6%
Average time per assignment	Average time per assignment	Percentage decrease in assessment time

Qualitative data from the teacher perception questionnaires and semi-structured interviews provided rich insights into the perceived benefits and challenges of using VLMs. A thematic analysis revealed the following key themes:

1. Increased Efficiency: Teachers consistently reported that VLMs significantly reduced the time required for assessment, allowing them to allocate more time to instructional activities. One teacher stated, "Using the VLM was like having an extra pair of hands. I could assess a whole stack of essays in the time it used to take me to do just a few."

2. Enhanced Objectivity: Teachers perceived VLMs as providing more objective and consistent assessments compared to traditional methods. They believed that VLMs helped to reduce the influence of personal biases in grading. As one participant put it, "The VLM applied the rubric consistently across all students, which is something that's hard to do when you're tired and grading a lot of papers."

3. Detailed Feedback: Teachers valued the ability of VLMs to provide detailed feedback to students, highlighting both strengths and areas for improvement. Some teachers noted that the feedback generated by the VLMs was more specific and actionable than the feedback they typically provided. "The VLM's feedback pointed out things I wouldn't have thought to mention, especially in terms of suggesting improvements," stated one participating teacher.

4. Implementation Challenges: Some teachers expressed concerns about the lack of transparency in the VLM assessment process and the need for adequate training and support to effectively use the technology. Teachers also stated that more training is needed to properly implement the prompts.

These findings suggest that VLMs hold promise for transforming student assessment practices. The quantitative data provides evidence of increased efficiency, while the qualitative data highlights perceived benefits related to objectivity and feedback quality. However, it is also important to acknowledge the implementation challenges reported by teachers. The next step would be to analyze VLMs in more schools and different subjects. This will give insight on the efficacy of VLMs in different assessment areas. These results confirm the potential of AI tools in transforming educational practices, a notion advocated by Lee, Cho, and Kim [8], while reinforcing the importance of proper teacher training for successful integration and meaningful outcomes

D.   Conclusion

This study provides compelling evidence for the potential of Vision Language Models (VLMs) to transform student assessment practices in the context of the Kurikulum Merdeka. Key findings demonstrate that VLMs can significantly enhance assessment efficiency, leading to substantial reductions in teacher workload, and improve the perceived objectivity and consistency of evaluations. The detailed feedback provided by VLMs offers a valuable resource for students, highlighting both their strengths and areas for improvement. The case study at SMP N 32 Padang demonstrates the real-world applicability of these technologies.

The implementation of VLMs is not without its challenges. Addressing concerns about transparency and providing adequate training and support for teachers are crucial for ensuring successful adoption. Future research should focus on developing more transparent and explainable VLM assessment models, as well as exploring strategies for integrating VLMs into diverse educational settings and subject areas. Further research is needed to explore how students benefit from the assessment and how the assessment can be better implemented. This is needed because although teachers found it effective, students' experience with the technology is not yet explored.

These findings suggest that VLMs represent a promising avenue for innovation in education, offering the potential to create more efficient, objective, and personalized learning experiences for students. Continued research and development in this area are essential to fully realize the transformative power of AI in education.

E.   References

[1]   Kemdikbudristek, "Kurikulum Merdeka: Pedoman umum," 2022.

[2]   W. J. Popham, *Classroom assessment: What teachers need to know*, 7th ed. Pearson, 2014.

[3]   S. M. Brookhart, *How to give effective feedback to your students*, 3rd ed. Association for Supervision and Curriculum Development, 2018.

[4]   D. Higgins, Y. Xiao, and K. Zechner, Eds., *Handbook of Automated Essay Evaluation: Current Applications and New Directions*. Routledge, 2016.

[5]   M. Chen, C. Leung, W. Xie, W. Ma, B. Li, and J. Zhang, "Deep learning for vision and language integration: A comprehensive review," *Artif. Intell. Rev.*, vol. 56, no. 4, pp. 2579–2625, 2023, doi: 10.1007/s10462-022-10231-3.

[6]   A. Radford *et al.*, "Learning transferable visual models from natural language supervision," *arXiv Prepr. arXiv2103.00020*, 2021.

[7]   V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77–101, 2006, doi: 10.1191/1478088706qp063oa.

[8]   J. Lee, Y. Cho, and J. Kim, "The impact of AI on education: A perspective from the educational technology research," *Educ. Technol. Res. Dev.*, vol. 69, no. 3, pp. 575–602, 2021, doi: 10.1007/s11423-021-09984-2.

[9]   Mulawarman Munsyir, S.E., S.SI., M.Kom, Harry Setya Hadi, S.Kom, M.Kom, Tata Sumitra, S.Kom., M.Kom, Ita Arfyanti, S. Kom, M. M, and Arnes Yuli Vandika, "Algoritma dan Pemrograman: Pendekatan Komprehensif", *YPAD*, vol. 1, no. 1, Jul. 2024.