UDC 004.05

FMMT Marikar, M Aman

INTEGRATING AI INTO PHYSICAL TEACHER ASSESSMENT: A SRI LANKAN CASE STUDY

Abstract Assessment practices, including individual and group marking, are subject to biases, such as gender and family status. This study explored the impact of these factors on marking consistency and accuracy. Data were collected by dividing answer scripts among five groups and analyzing one set using AI tools (Bard and ChatGPT). Results revealed significant differences in marking, with AI-generated scores being lower than human-assigned scores. This highlights the limitations of AI in capturing nuanced understanding. To address these issues, a combination of human and AI assessment, along with robust marking schemes, can be employed to improve the fairness and accuracy of evaluations.

Keywords: Assessment, Marking, Bias, AI, Human Judgment, Consistency, Accuracy

Introduction

Ensuring fairness in examinations is a complex challenge, influenced by various factors such as personal biases, geographical disparities, and external pressures. Subtle forms of unfairness, like time constraints and resource limitations, can significantly impact vulnerable communities, leading to compromised marking quality [1]. While conference marking can mitigate bias, it cannot completely eliminate it. Independent marking, without a robust mechanism, poses risks of inconsistent and potentially unfair assessments [2]. Therefore, developing effective marking procedures that balance standardization with flexibility is crucial to ensure fair and equitable evaluations, particularly for marginalized groups.

AI-powered automated grading systems can efficiently analyze and assess large volumes of student work, particularly for multiple-choice and short-answer questions [3]. These systems can quickly provide accurate and consistent feedback, freeing up educators' time for more complex tasks. For instance, platforms like Gradescope and Turnitin utilize AI to streamline the grading process, ensuring fairness and accuracy [4].

AI-driven personalized learning platforms can tailor educational content and assessments to individual student needs. By analyzing student performance data, these platforms can identify knowledge gaps and recommend targeted learning resources [5, 6]. For example, platforms like Knewton and Khan Academy use AI to create personalized learning paths, helping students to learn at their own pace and do the assessment [7].

AI-powered exam proctoring tools can monitor students remotely during online exams, reducing the risk of cheating. These tools can detect suspicious behavior, such as

multiple login attempts or unusual eye movements, and alert proctors to potential irregularities. Proctorio and ExamSoft are examples of AI-powered proctoring solutions that enhance exam security [8, 9].

When assessing an individual's capacity to integrate AI tools into examination processes, a comprehensive evaluation of their knowledge, skills, and attitudes is essential. This assessment should encompass their understanding of AI concepts, their proficiency in using AI tools effectively, and their ability to critically evaluate AI-generated outputs [10]. Additionally, their awareness of relevant policies and ethical considerations, as well as their willingness to embrace change and adopt new technologies, are crucial factors to consider. By evaluating these dimensions, we can gain insights into an individual's readiness to leverage AI tools to enhance the quality and fairness of examinations.

Human expert assessment would involve experienced Sri Lankan physical education teachers to validate and refine AI-generated assessments. Peer review among teachers could also be encouraged to provide constructive feedback. By combining the quantitative data from AI analysis with qualitative insights from human experts, personalized feedback reports could be generated for Sri Lankan physical education teachers, highlighting strengths and areas for improvement. Continuous learning and data-driven decision-making would be crucial to ensure the ongoing improvement of AI-based assessment tools and their impact on teacher performance and student outcomes in Sri Lanka.

Methodology

To integrate AI into physical teacher assessment in Sri Lanka, a comprehensive approach involving data collection, AI-powered analysis, human expert assessment, and iterative improvement is necessary. Data collection would involve video analysis of physical education classes conducted in Sri Lankan schools, capturing teacher-student interactions, skill demonstrations, and class dynamics. Sensor data, such as heart rate and movement patterns of teachers and students, could be collected using wearable devices. Student feedback can be gathered through surveys and interviews.

AI-powered analysis would employ computer vision to analyze video footage, identifying teacher's movement patterns, posture, and facial expressions. Machine learning algorithms could be used to predict student performance based on various factors, including teacher behavior, class dynamics, and student engagement. Natural language processing could be employed to analyze student feedback and identify key themes and suggestions for improvement.

Questionnaire Used: Part of the mock examination answer script was given as a questionnaire to be analyzed by the lectures and using AI tool it is as follows.

“It is very difficult to get breast milk contaminated. Contamination of baby food leads to infection and poor growth. The final thing can be a weak nation. Europe’s women

 ISSN 1560-8956

don’t like breast feeding. They say it can change the shape of the breast. No evidence for this situation. It does not drop with breast feeding mothers like breast feeding. They feel happy when child drinks. The child also feels happy. Multinational companies make mothers stop breast feeding. They give free milk to mothers to stop them their own breast milk. This was done to for them to get more money. We should not be fooled by propagandas. Last year American University students boycotted a big American firm that makes infant food same reason. If American people can protest against their own company, why we cannot do like that? We should encourage our mothers to breast feed as long as possible. That is why if they do like that lotof money can be used for other things and our country can benefit. Breast milk is the best milk. When we discuss the statement, we have to remember that mothers have been feeding their infants which own milk for ages. The west influence changed the practice in this country to last few years and it in the good that we go back to what those we have been used to a long time.

Breast milk was the best composition to suit the needs of humans. As the saying goes breast milk is best for kid which cow’s milk is best for calves his requirements of the infants can be easily met if breast milk is used.

A big problem in an infant is infections. Breast milk can counteract infection in childhood because pressure if gamma globulins. Many children end up in problems due to infections in childhood we prevent them with breast milk.

Recent findings in American shows that breast milk promotes better growth. Better growth in infancy means better growth citizen. The optimum temperature of breast milk of human is also important. The baby has same temperature like mother and he wishes to drink it.”

Administration of Questionnaire among the Students and AI tool. The hard copy of answer script which is used as a questionnaire was administered among each of the respective participants at the University teacher training programme. Given 30 minutes and within given period they have to mark the answer script. Care was taken to avoid exchanging the participants ideas. One set was given to the ChatGPT and Bard to check and give the score.

Data Analysis To analyze the marks of the answer script, we compared informal reasoning displayed by individuals representing the highland low level of marking with the standard deviation and it was furhter analyzed by the Chat GPT and Bard for comparison. The validity of the marking was independently assessed by two observers. For statistical analysis, we transformed all our data using the basic statistical analysis package.

Results and discussion

Table 1.

Frequency analysis of group marking – Physical assessment

	Group1	**Group2**	**Group3**	**Group4**
Mean	55.40	49.25	45.40	56.60
Median	55.00	49.50	40.00	56.00
Mode	48	37	40	49
Std. Deviation	5.320	10.012	13.221	6.693
Range	13	24	34	17
Minimum	48	37	34	49
Maximum	61	61	68	66
Average	55	49	45	56

The mean values for Group 1, Group 2, Group 3, and Group 4 were observed to be 55.4, 49.25, 45.40, and 56.00, respectively. It was noted that Group 3 had a significantly higher standard deviation of 13.22 compared to the other groups. This high standard deviation in Group 3 was likely due to the diverse backgrounds of its members, as they were not from the same discipline or gender. The range of marking in Group 3 was also observed to be significantly high, with a value of 34.

Outcome of AI tool (Bard and ChatGPT)

Importance of Breastfeeding vs. Formula Feeding: Addressing Misconceptions (Score: 30/100)

Your text contains some misconceptions about breastfeeding and formula feeding. Here's a breakdown:

Breastfeeding Advantages (Accurate):

- *Best source of nutrients:* Breast milk is ideally suited for a baby's needs, containing essential vitamins, minerals,and antibodies.
- *Promotes immune health:* Breast milk contains antibodies that help protect babies from infections.
- *Emotional bonding:* Breastfeeding can promote a strong bond between mother and baby.
- *Convenience and cost-effective***:** Breastfeeding requires no preparation and is free.

*Misconceptions***:**

- *Formula contamination***:** Both breast milk and formula can become contaminated if not handled properly. Hygiene is crucial.
- *Breast shape:* There's no scientific evidence that breastfeeding permanently changes breast shape.
- *European women and breastfeeding:* While breastfeeding rates may vary, there's no general dislike in Europe.

 ISSN 1560-8956

- *Formula companies and propaganda:* While some marketing practices may raise concerns, reputable companies follow regulations.
- *Week nation (likely meant "weak nation"):* Breastfeeding doesn't determine national strength.
- *Teeth at birth:* Babies don't have teeth because their jaws aren't developed enough for chewing, not for breastfeeding.

Formula's Role:

- *Viable alternative:* Formula is a safe and healthy option for babies who cannot breastfeed.
- *Supplementation:* Formula can supplement breastfeeding if milk supply is insufficient.

Table 2.

Marking schemes – Physical assessment

Group one Marking Scheme	Group two Marking Scheme	Group three Marking Scheme	Group four Marking Scheme
Introduction (20 Marks) about breast milk Sri Lankan situation Foreign situation	Organization and Flow (20 Marks)	Introduction (20 Marks)	Content (60 Marks) Relevance (10) Analysis (10) Adequate of facts (30) Accuracy of facts (10)
Importance of brest feeding (50 Marks) protect against infection Provide Vitamins/Ca best milk	Language and Grammer(20 Marks)	Problem Statement (20 Marks)	Language (20 Marks)
Research report (10 Marks) Present past	Introduction (15 Marks)	Purpose (20 Mariks)	Presentation (20 Marks)
Present threat (5 Marks)	Milk companies provides only milk Importance (15 Marks)	Significance (10 Marks)	
Actions against these importers (5 Marks) USA/China	If it is not (15 Marks)	Statistics (20 Marks)	
Our duty (10 Marks)	Present role of promoting Breast Feeding Conclusion (15 Marks)	Style (10 marks)	

According to Table 2, four different types of marking schemes were observed. It was observed that two deviations occurred in the second marking. Comprehensive marking

schemes were provided by Groups One and Three, resulting in the highest mean value of 56. Groups Two and Three allocated the highest portion to unnecessary sections, leading to errors in their marking.

AI conclusion remarks:

Breastfeeding is generally recommended for its numerous benefits. However, formula is a valuable alternative when breastfeeding isn't possible. The focus should be on feeding your baby the way that works best for you and your baby's health.

- 2: All information accurate.
- 1: Some information accurate, some minor misconceptions.
- 0: Mostly inaccurate information.

Accuracy of Formula Information | 20% | |

- 2: Formula recognized as a viable alternative/supplement.
- 1: Formula mentioned but role unclear or negative portrayal.
- 0: No mention of formula or inaccurate portrayal

Neutrality/Objectivity | 20% | |

- 2: Tone is neutral and avoids promoting one over the other unfairly.
- 1: Leans towards breastfeeding but acknowledges formula as an option.
- 0: Strongly promotes breastfeeding and demonizes formula.

Clarity and Conciseness | 15% | |

- 2: Information is easy to understand and avoids unnecessary details.
- 1: Information is clear but could be presented more concisely.
- 0: Information is confusing or poorly organized.

Sources/References (if provided) | 15% | |

- 2: Includes credible sources (WHO, CDC) to support claims.
- 1: Mentions sources but not credible or not relevant.
- 0: No sources provided.

Total Score (out of 100):

Grading:

- 80-100: Excellent - Accurate and objective information on both breastfeeding and formula.
- 60-79: Good - Mostly accurate information with minor issues.
- 40-59: Fair - Needs improvement in accuracy and/or objectivity.
- Below 40: Poor - Misleading or inaccurate information.

Additional Notes:

- This is a sample marking scheme, and specific weights can be adjusted based on the context.
- The scheme encourages factual information and avoids emotional appeals.

 ISSN 1560-8956

Remember, the goal is to provide clear and balanced information for parents to make informed decisions regarding infant feeding. The integration of AI into the marking process offers a promising solution to the challenges of human bias and inconsistency. By analyzing vast datasets of graded assignments, AI algorithms can identify patterns and develop objective criteria for assessment. This can significantly improve the accuracy and fairness of the marking process.

A hybrid approach that combines human expertise with AI can lead to the best possible outcomes. AI can automate routine tasks such as grading multiple-choice questions and providing initial feedback on short answer responses, freeing up time for human markers to focus on more complex tasks like evaluating essays and providing in-depth feedback. Additionally, AI can help identify potential biases in marking schemes and suggest improvements to ensure fairness.

Conclusions

By leveraging the power of AI, educators can create a more efficient, equitable, and reliable marking process. This can lead to improved student learning outcomes and a more effective educational system overall. As AI technology continues to advance, we can expect to see even more innovative applications in the field of education, ultimately transforming the way we teach and learn.

REFERENCE

1. *Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J.M., Milligan, S., Selwyn, N. and Gašević, D.*, 2022. Assessment in the age of artificial intelligence. Computers and Education: Artificial Intelligence, 3, p.100075. https://doi.org/10.1016/j. caeai.2022.100075

2. *Esarey, J. and Valdes, N.*, 2020. Unbiased, reliable, and valid student evaluations can still be unfair. Assessment & Evaluation in Higher Education, 45(8), pp.1106-1120. https://doi.org/10.1080/02602938.2020.1724875

3. *Oc, Y. and Hassen, H.,* 2024. Comparing the effectiveness of multiple-answer and single-answer multiple-choice questions in assessing student learning. Marketing Education Review, pp.1-14. https://doi.org/10.1080/10528008.2024.2417106

4. *Gehringer, E.F., Menon, A. and Wang, G.*, 2021, July. Tools for detecting plagiarism in online exams. In 2021 ASEE Virtual Annual Conference Content Access. https://doi.org/10.18260/1-2--37915

5. *Li, Y., Jiang, Y., Li, Z. and Xia, S.T.*, 2022. Backdoor learning: A survey. IEEE Transactions on Neural Networks and Learning Systems, 35(1), pp.5-22. https://doi.org/ 10.1109/TNNLS.2022.3182979

6. *Fernando, S.Y. and Marikar, F.M.*, 2017. Constructivist teaching/learning theory and participatory teaching methods. Journal of Curriculum and Teaching, 6(1), pp.110-122.

7. *Ip, K.*, 2024. The rise of EdTech: Transforming education through entrepreneurial ventures. Advances in Online Education: A Peer-Reviewed Journal, 3(2), pp.177-193. https://doi.org/10.69554/NSVD4541

8. *Nagpal, N., Srivastava, A. and Verma, P.*, 2024. AI-Powered Proctoring: Safeguarding Online Assessment in the Education 5.0. In Explainable AI for Education: Recent Trends and Challenges (pp. 271-285). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72410-7_15

9. *Marikar, F.M.*, 2018. Gender stereotypes influence on assessment in a multidisciplinary platform. International Journal of Educational Studies, 5(1), pp.43-47.

10. *Zybaczynska, J., Norris, M., Modi, S., Brennan, J., Jhaveri, P., Craig, T.J. and Al-Shaikhly, T.*, 2024. Artificial Intelligence–Generated Scientific Literature: A Critical Appraisal. The Journal of Allergy and Clinical Immunology: In Practice, 12(1), pp.106-110. https://doi.org/10.1016/j.jaip.2023.10.010