



**National Conference on Educational Measurement and Evaluation (NCEME)**

**2024**

August 29-31, 2024

De La Salle University - Manila



# Ability of ChatGPT4 to Generate Standardized Assessment in Physics 1

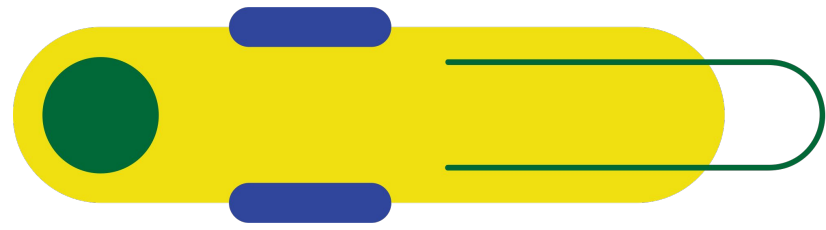
**Ms. Isalyn F. Camungol, MST-PHY**

**Colegio De Muntinlupa**

**Lydia S. Roleda, Ph.D.**

**DLSU-Manila**

**August 30, 2024**



# Ability of ChatGPT4 to Generate Standardized Assessment in Physics 1



**Background of the Study**



**Research Questions**



**Methodology**

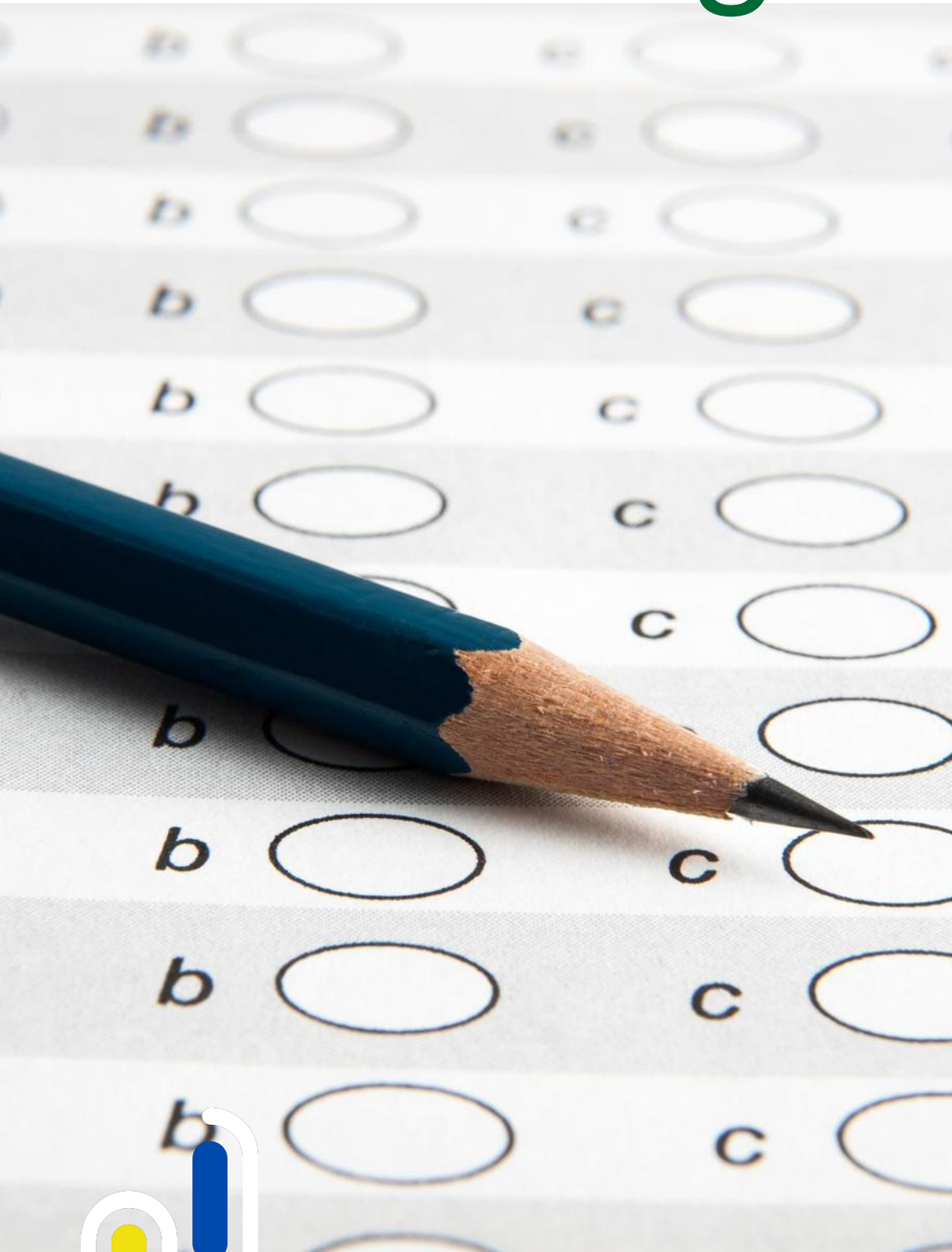


**Analysis**



**Conclusions and  
Recommendations**

# Background of the Study



**Standardized assessments** play a critical role in higher education institutions by providing a consistent and objective measure of student performance across different institutions and programs. These assessments enable educators and administrators to evaluate the effectiveness of their curricula and identify areas where students may need additional support or resources. Standardized test items offer a way to gauge the educational attainment of students uniformly, thereby facilitating comparisons and benchmarking across various educational settings (Popham, 2010).



# Background of the Study



- The emergence of Artificial Intelligence in Education (AIED) marks a transformative shift in how educational systems operate and evolve. Early research in this field focused on the potential of AI to enhance instructional methods and learning outcomes.

# Background of the Study



◦At the forefront of the AI-driven system is ChatGPT, an acronym for Chat Generative Pre-trained Transformer, representing a prominent CHATBOT model created by OpenAI. It was officially introduced on November 30, 2022, and has gained widespread recognition, boasting more than 100 million users in just five days—setting it apart from other AI-driven chatbots (OpenAI, 2023) and continues to grow exponentially, which is set to reach 514 million users and 1.8 billion visits per month.

◦Chatbot technology incorporates generic language models extracted from large parts of the Internet and enables feedback by limiting themselves to text or voice interfaces. Research on AI employs ChatGPT as an example of a sophisticated chatbot that can be used to explore the effects of chatbots in education (Tlili et al., 2023) (Zhai, 2023) (Baidoo-Anu and Ansah, 2023).

◦ChatGPT as a powerful chatbot exhibits the ability to engage in meaningful, human-like conversations, provide detailed explanations, and offer subject-specific guidance. It stands as a potential virtual companion or can be a personal tutor in the student's educational journey, with the capacity to answer questions, clarify concepts, and stimulate critical thinking (West,

2022) (Bandeau et al., 2022)



# Background of the Study



- personal virtual study guide
- instant feedback for the students queries
- the application of ChatGPT in formative assessment practices within mathematics education is gaining attention. Research suggests that ChatGPT can be utilized to generate a wide variety of practice problems and quizzes, instantly grading them and providing detailed feedback.

# Background of the Study



- personal virtual study guide
- instant feedback for the students queries
- the application of ChatGPT in formative assessment practices within mathematics education is gaining attention. Research suggests that ChatGPT can be utilized to generate a wide variety of practice problems and quizzes, instantly grading them and providing detailed feedback.

## Scope of the Study

1) For selected Physics topics only (Scalars and Vectors, Mechanics, Kinematics, Dynamics and Fluid Dynamics)

2) Used the ChatGPT4 only

## Relevance of the Study

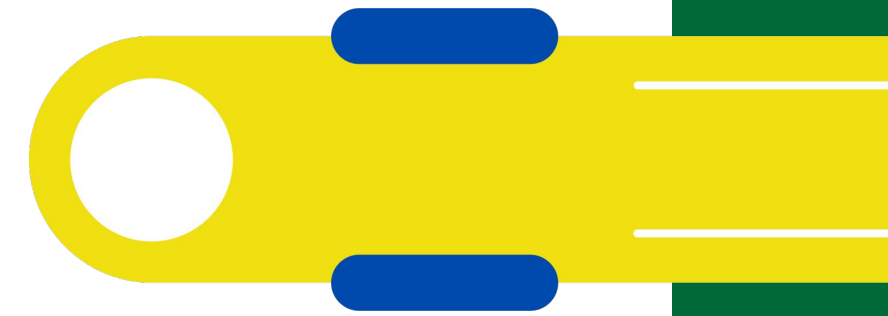
- Preliminary study on leveraging the use of AI Chatbots in education, specifically in creating valid and reliable assessments and Physics Education.

## Research Question

1. ) What are the difficulty indices and discriminatory indices of Physics multiple-choice items generated from ChatGPT-4?
2. ) Is ChatGPT4 can be a potential tool in creating standardized assessment?



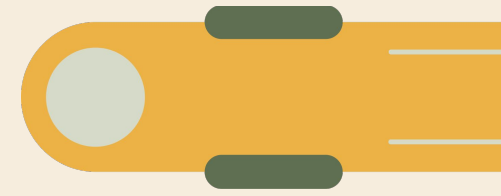
# Methodology



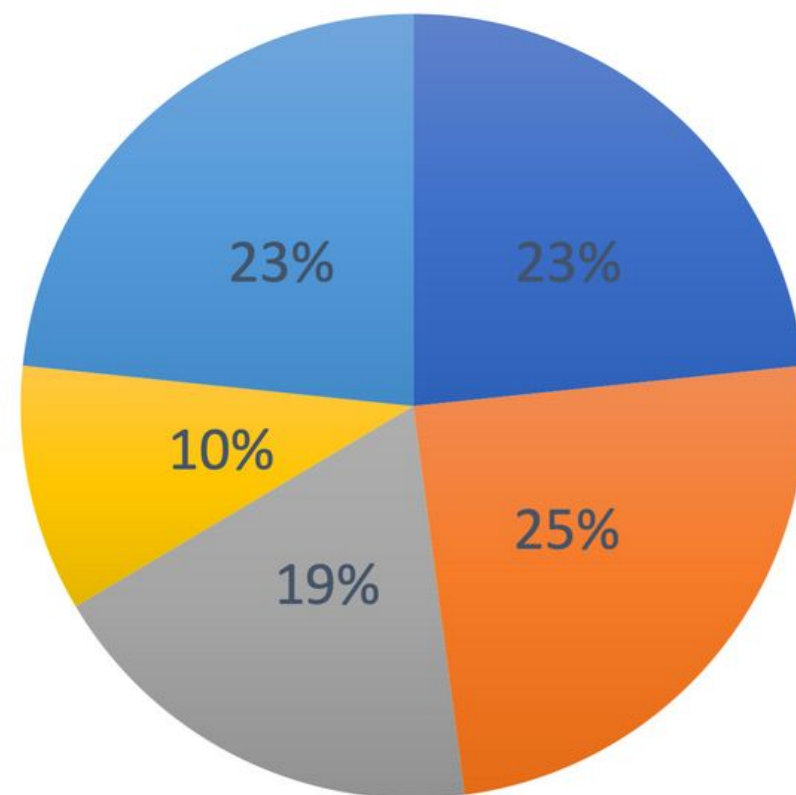
## **Quantitative Methods**

Measure and analyze the descriptive statistics of the difficulty indices and discrimination indices of the generated items from ChatGPT4.

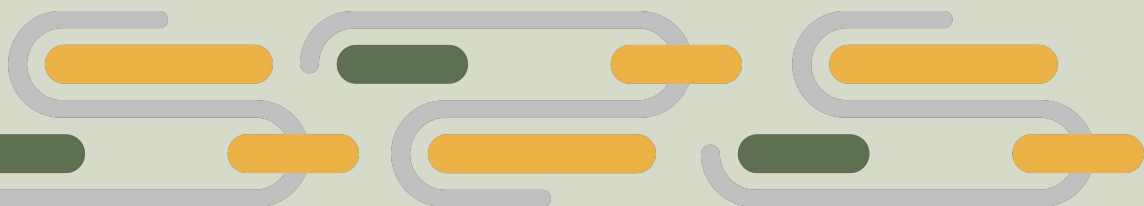
# Quantitative Data



Percentage of the Quality of Items  
Generated from ChatGPT4



Number of Items	Quality
39	Excellent (Good P-Score and DI)
41	Good (Acceptable P-Score and DI)
31	Easy (High P-Score and Acceptable DI)
17	Difficult (Low P-Score and Acceptable DI)
39	Rejected
<b>167</b>	



# Analysis

Analyzing the quality of the items based on their discriminatory factors and p-scores, 23.08% of the items are of the best quality, indicating highly effective discrimination and moderate difficulty. 24.26% of the items are of good quality, showing good discrimination and good discriminatory indices. 28.40% of the items have acceptable discrimination but are either too easy or too difficult for the students. However, 23.08% should be rejected.



# conclusions and recommendations

- This result indicates that educators can harness the power of AI to create good quality assessments in each discipline without consuming so much time and even creating a standardized assessment.
- The analysis also highlights areas for improvement. Some items generated by ChatGPT are either too easy or too difficult, which may not contribute effectively to the overall assessment, indicating the need for human-AI collaboration.



Thank You