

DESIGNING REAL-WORLD STATISTICS PROBLEMS ON CENTRAL TENDENCY OF UNGROUPED DATA USING CHATGPT 3.5

Truong Huu Hung¹ and Le Thuy Dieu²

¹*Chu Van An Lower Secondary School, Hanoi city, Vietnam*

²*Faculty of Mathematics and Information Technology, Hanoi National University of Education,
Hanoi city, Vietnam*

*Corresponding author: Truong Huu Hung, e-mail: truonghuuhung2608@gmail.com

Received June 20, 2024. Revised October 20, 2024. Accepted December 27, 2024.

Abstract. Statistics - Probability stands as one of the three fundamental strands within the new Vietnamese Mathematics General Education Curriculum (2018). This development significantly expands the scope of teaching and learning opportunities in the realm of practical mathematics as the application of Statistics in everyday life is an ever-present reality that touches various facets of society, especially in the contemporary data-driven landscape. The birth of ChatGPT has conferred numerous benefits upon educators, one of which is streamlining the lesson preparation process. Through a structured procedure, real-world problem design is guided by a series of techniques aimed at enhancing question quality. The research method involves the iterative refinement of questions and their alignment with cognitive learning levels, culminating in a mini-test created using ChatGPT. This structured approach demonstrates the potential for AI to aid in educational content development, particularly in the domain of statistical problem generation.

Keywords: Statistics, Grade 10, central tendency of ungrouped data, ChatGPT, real-world mathematics problems.

1. Introduction

Mathematics, as Boyer and Merzbach (2011) and Schubring (2014) assert, is a practical science deeply ingrained in real-life applications [1], [2]. Its relevance extends beyond the concepts of arithmetic, algebra, or geometry, spreading through fields like finance, economics, medicine, and sociology [3]. To engage students in mathematics, it's important to show how it can be useful and relevant in everyday life.

In alignment with this argument, the Statistics-probability strand in the Mathematics General Education Curriculum emphasizes applying statistical rules in practical scenarios [4]. Neumann, Hood, and Neumann (2013) advocate for significant shifts in pedagogical methodologies, advocating for less theory and more data-centric learning, fostering statistical thinking, and integrating contextual applications to enhance statistical inference [5].

However, challenges persist, particularly in the classroom setting. Andriani and Fauzan (2019) highlight students' struggles with statistical concepts like mean, mode, and median, attributing these difficulties to the disconnect between theoretical learning and practical

application. The lack of practical connections within mathematical teachings contributes to the challenges faced by students [6].

Gil and Ben-Zvi (2011) emphasize that it is important not only to understand statistics but also to connect it to real-world situations [7]. Hence, connecting theoretical knowledge with practical applications becomes imperative for effective learning.

In order to streamline learning and teaching processes, Das et al. (2021) emphasize the potential of Automatic Question Generation (AQG), which is powered by Artificial Intelligence (AI) [8]. AQG, when integrated with technologies such as ChatGPT, not only lessens the manual workload for educators but also enhances scalability in creating contextually relevant and high-quality questions [9].

Our paper studies the combination of real-world problems and AI-driven question generation through ChatGPT 3.5. This paper answered the question: “How can ChatGPT 3.5 be used to design real-world statistics problems on the central tendency of ungrouped data that meet the requirements of the Grade 10 Vietnamese Mathematics General Education Curriculum (2018), and what techniques can be applied to enhance the quality of these problems?” This synthesis advocates for a paradigm shift in pedagogy, emphasizing the importance of contextualized, practical, and engaging approaches to Statistics instruction.

2. Content

2.1. The requirements in 10th Grade Statistics content knowledge

The 2018 Vietnamese General Education Curriculum aims to enhance students’ competencies and qualities, with a focus on developing practical problem-solving skills [10]. In high schools, the requirements of the Statistics - Probability strand include students being able to (1) collect, classify, perform, analyze, and process statistical data; (2) use statistical data analysis tools through measures of central tendency and measures of dispersion for ungrouped and grouped data samples; and (3) use statistical rules in practice [4]. In particular, the requirements for Grade 10 Statistics content knowledge are shown in Table 1 [4].

Table 1. The requirements of Grade 10 Statistics content knowledge

Content	Requirements
The measures of central tendency for ungrouped data	- Calculate the measures of central tendency for ungrouped data: mean, median, quartiles, and mode. - Explain the meaning and role of these measures in realistic sample data. - Indicate the conclusions based on the meaning of the measures of central tendency for ungrouped data in simple cases.
The measures of dispersion for ungrouped data	- Calculate the measures of dispersion for ungrouped data: range of variation, range of interquartile, variance, and standard deviation. - Explain the meaning and role of these measures in realistic sample data. - Indicate the conclusions based on the meaning of the measures of dispersion for ungrouped data in simple cases. - Recognize the relationship between Statistics and other disciplines in the Grade 10 curriculum and real life.

In this paper, the content of measures of central tendency for ungrouped data is our focus. The requirements of this content involve using Bloom's Taxonomy at various levels of learning. To be more specific, the first requirement asks students to apply their knowledge of statistical formulas and procedures to calculate the measures of central tendency for ungrouped data, placing it at the "Apply" level of Bloom's Taxonomy. At the "Understand" level, students need to comprehend why these measures are used, what information they provide about the data, and how they are applied in real-world scenarios, which are stated in the second requirement. Finally, drawing conclusions and making inferences based on the analysis of measures of central tendency aligns with the "Analyze" level of Bloom's Taxonomy.

These requirements collectively provide a well-rounded approach to learning measures of central tendency for ungrouped data, from calculations to understanding the meaning and implications of these measures in real-world contexts.

2.2. Real-world statistics problems

2.2.1. Real-world mathematics problems

A recurring challenge in mathematics education is the lack of perceived relevance among students, triggering the common query, "why do we have to learn this?" [11]. This disconnection between mathematics and real-life applications underscores the need to establish meaningful connections between the two. Word problems serve as a typical bridge to bring theory-heavy subjects like mathematics closer to real-life contexts [12].

Policy documents across various countries emphasize the importance of integrating real-world applications into mathematics education, which implies a need for curricula to focus on practical, real-life scenarios to enhance the relevance of mathematics for students [13].

Realistic Mathematics Education (RME) emphasizes the concordance between tasks with real-life contexts and students' abilities to engage with and model these situations [1]. RME acknowledges that the context of a problem should be relatable and experiential for students, whether derived from real-life experiences, fairy tales, or the formal world of mathematics [1].

2.2.2. Real-world statistics problems

According to Moore and Cobb (1997), a different kind of thinking is required for Statistics, as numbers in data possess a context, not just mere numbers [2]. By understanding the context, students can make connections to the real world and reach conclusions that are both relevant and meaningful.

Real-world Statistics problems can serve as an approach to assess students' statistical knowledge. Unlike assessments that focus solely on theoretical knowledge, these problems require students to apply their statistical skills to real-life situations. Moreover, real-world contexts might help engage students and foster a sense of curiosity, motivating them to delve deeper into statistical concepts.

Crafting problems that align with both curriculum objectives and real-world relevance requires time and effort. ChatGPT can assist educators in generating diverse and challenging real-world Statistics problems. It leverages vast datasets to simulate real scenarios, providing educators with a wide array of problem types that enhance the student's learning experiences.

2.3. Designing real-world statistics problems using ChatGPT 3.5

2.3.1. A proposed procedure

A structured approach can guide the process of developing real-world statistics problems using ChatGPT. This enables a more focused and organized creation process, ensuring that the generated problems meet the desired criteria in terms of relevance, complexity, and suitability for educational purposes [14].

** Step 1. Create an academic environment*

To initiate the process, it is essential to create an academic environment for ChatGPT (Figure 1). This entails assigning ChatGPT a specific role within the context of the Statistics problem. This helps streamline the communication process with ChatGPT and allows the AI model to better comprehend the user's expectations, guiding its actions or responses towards fulfilling the designated role effectively.

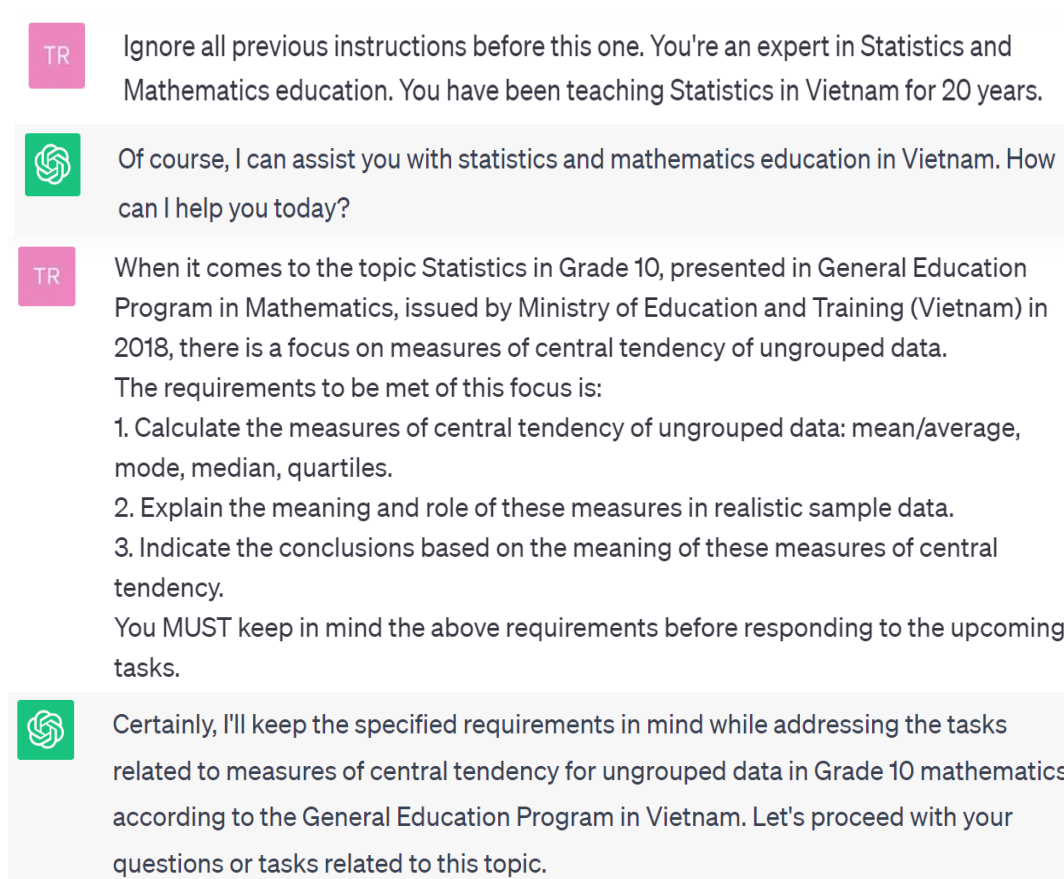


Figure 1. Creating an academic environment

** Step 2. Make a request*

Once the academic environment is established, the next step is to make a well-defined request to ChatGPT (see Figure 2). The more detailed and specific the request, the better the generated Statistics problems are likely to be.



You

I am delivering a lesson of Statistics for 10th grade students. Propose a question satisfying ALL of the following conditions:

- It adheres to the first requirement.
- It includes 4 alternative answers, 3 incorrect and 1 correct.
- Question completion time: 1 minute.
- Use passage stimulus for the question. The stimulus MUST relate to Vietnamese context.
- Include data shown in table.



ChatGPT

Certainly, here's a multiple-choice question that meets all of your specified conditions:

Question:

In a survey conducted in a Vietnamese city, the following table shows the number of hours per week that people spend on various leisure activities.

Leisure Activity	Number of Hours per Week
Watching TV	10
Playing Sports	8
Reading Books	5
Social Media	15
Total Hours Spent	38

What is the mean (average) number of hours per week that people spend on leisure activities in this city?

- A) 12 hours per week
- B) 7.5 hours per week
- C) 9.5 hours per week
- D) 6 hours per week

Figure 2. Making a request

By providing ChatGPT with precise guidelines and instructions, the problems it generates can align closely with the educational objectives (what requirements to be followed) and desired level of complexity (what types of questions to be used, how long it takes to complete the question, in what form the questions are presented, etc.). When creating real-world problems, a helpful method is to use textual stimuli. Passage stimuli serve as a valuable tool for contextualizing statistical problems within a descriptive framework.

2.3.2. Additional tasks done during the process

In the process of designing real-world Statistics problems with ChatGPT, several additional tasks are essential to enhance the quality and diversity of the questions generated.

*** Task 1. Create multiple questions with different levels**

One crucial task is to create multiple questions with varying levels of complexity (Figure 3). This can accommodate a wide variety of students, from those seeking basic comprehension to those aiming for more advanced problem-solving skills. This diversity ensures that the generated problems can be adapted to different educational contexts.

- TR** Propose questions satisfying ALL of the following conditions:
- They adhere to the third requirement.
 - They are all open-ended questions.
 - They include all levels of difficulty from C1 to C6 using Bloom's Taxonomy.
 - Each question is used with different passage stimulus. The stimulus MUST relate to Vietnamese context.
 - Three questions include real data shown in table.

Figure 3. A prompt for creating multiple questions with different levels

Almost all the questions generated meet all the conditions. For example, the sixth question is a relatively comprehensive one (see Figure 4).

- First, it involves interpreting profit margin data, employing measures of central tendency, and drawing conclusions about business profitability, aligning with the understanding of measures of central tendency and what conclusions they can indicate.
- Secondly, the question demands the creation of a strategy, that involves higher-order thinking, innovation, and the synthesis of financial analysis skills to generate new solutions tailored to the Vietnamese business landscape, which is suitable for the level of "Create" in Bloom's Taxonomy.
- Thirdly, the question prompts the creation of a comprehensive strategy, allowing for diverse and innovative approaches to assessing and improving profitability, encouraging a full answer rather than seeking a single correct answer.
- Fourthly, although not explicitly outlined in the question, contextualizing the strategies within the Vietnamese business environment could involve considering factors like local market conditions or economic policies to ensure the relevance and applicability of the devised strategies to the Vietnamese context.
- Finally, the question includes data shown in the table.

Stimulus 6 - Vietnamese Business Profitability

Passage: Business profitability is a critical factor in Vietnam's economy.

Table: Quarterly Profit Margins (in %) of Different Businesses in Vietnam

Business	Q1 Profit Margin	Q2 Profit Margin	Q3 Profit Margin	Q4 Profit Margin
Retail Chain	7.2	7.5	7.4	7.6
Tech Startup	12.1	12.5	12.2	12.3
Manufacturing	9.5	9.6	9.4	9.7

C6: Question for Creation (Evaluation)

- As a financial analyst in Vietnam, outline a comprehensive strategy for assessing and improving the profitability of businesses using the provided profit margin data.

Figure 4. The 6th proposed question of the different-level questions generation process

However, during this process, there are common mistakes that can be encountered. One such mistake is using a single stimulus for several questions, resulting in a lack of the necessary diversity. Another issue is the potential shortage or redundancy of questions. Furthermore, there may be instances where the system produces more or fewer than three questions with data presented in a table. This inconsistency can disrupt the intended structure of the problem set. In some cases, ChatGPT might generate questions that are unrelated to Statistics, which can be confusing and counterproductive for educational purposes.

To address these challenges, a viable solution is to provide feedback to ChatGPT and refine the generated questions until the desired result is achieved. Additionally, reducing the complexity of the initial requirements can lead to more accurate and relevant question generation.

*** Task 2. Extract and improve questions**

Another valuable task is the extraction and improvement of questions. ChatGPT can remember previous questions (see Figure 5) to continue the conversation in a more fluid manner.

Stimulus 2 - Vietnamese Traffic Congestion

Passage: Traffic congestion is a common issue in many Vietnamese cities, affecting daily commutes.

C2: Question for Application (Comprehension)

- Discuss how measures of central tendency, such as mean and median, can be used to analyze and address traffic congestion issues in Vietnamese cities.

Figure 5. The 2nd proposed question of the different-level questions generation process

To illustrate, ChatGPT can help identify and refine existing questions generated in the previous process, making them more suitable for teaching and assessment (see Figure 6).



You

Refine C2 questions with more information.

Stimulus 2 - Vietnamese Traffic Congestion

Passage: Traffic congestion is a common issue in many Vietnamese cities, affecting daily commutes. A recent study collected extensive data on traffic congestion during different times of the day in a major Vietnamese city, including average congestion indices, traffic flow rates, and accident reports.

C2: Question for Application (Comprehension) with Emphasis on Statistics

- How can the statistical data collected, including average congestion indices, traffic flow rates, and accident reports, be effectively analyzed and utilized to address traffic congestion issues in Vietnamese cities? Provide practical suggestions for city planners and commuters while considering the statistical insights derived from this data.

Figure 6. The edition of the 2nd proposed question

The original passage provides a general overview of traffic congestion without specific details. In contrast, the edited passage offers a clearer and more precise depiction by specifying the types of data gathered, such as average congestion indices and traffic flow rates. This specificity can help in understanding the statistical aspects of traffic congestion, such as how congestion indices are calculated, the relationship between flow rates and congestion, and the impact of accidents on traffic patterns.

*** Task 3. Conversion between question types**

ChatGPT can assist in transforming questions from one format to another, broadening the scope of learning and assessment. This can challenge students to apply their knowledge in different ways. By taking advantage of the continuous conversation with ChatGPT, a specific question generated before (see Figure 7) is now changed to different types (see Figure 8 and Figure 9).

Stimulus 4 - Vietnamese Coffee Consumption

Passage: Coffee is a popular beverage in Vietnam, and understanding consumption trends is essential for the coffee industry.

Table: Average Monthly Coffee Consumption (in cups) in Different Age Groups in Vietnam

Age Group	Mean	Mode	Median
Teens (13-19)	2.5	3	2
Adults (20-40)	3.2	3	3
Seniors (60+)	2.0	2	2

C4: Question for Synthesis (Analysis)

- Analyze the data on coffee consumption in different age groups in Vietnam and discuss the implications of these measures of central tendency for coffee marketers targeting these demographics.

Figure 7. The 4th proposed question of the different-level questions generation process

TR

You

Change C4 question into the following types:

- Multiple-choice question.
- Filling-gap question.
- Closed question.
- Multi-select question.
- Drag-and-drop question.
- Project-based question.

Figure 8. A prompt for conversion between question types

Multi-Select Question:

- Select all age groups that have a mean monthly coffee consumption exceeding 2.5 cups in Vietnam.
 - A) Teens (13-19)
 - B) Adults (20-40)
 - C) Seniors (60+)

Drag-and-Drop Question:

- Match the correct statistical measure to its description:
 - Drag "Mean" and match it with "Average monthly coffee consumption."
 - Drag "Mode" and match it with "Most frequently consumed amount of coffee."
 - Drag "Median" and match it with "Middle value of monthly coffee consumption."

Project-Based Question:

- Assume you are a marketing analyst for a coffee company operating in Vietnam. You have access to the provided data on coffee consumption among different age groups. Design a marketing campaign tailored to each age group based on their coffee consumption habits. Include strategies for promoting your coffee products and engaging with customers in each demographic.

Figure 9. Several different types are presented

The transformation of the original question into different types exhibits different approaches to evaluating understanding, encompassing both factual knowledge (closed question, multi-select question, drag-and-drop question, etc.) and higher-order thinking skills (project-based question, etc.). Furthermore, different question types cater to diverse learning styles and preferences among students, thereby ensuring engagement across a broader spectrum of learners. Exposure to different question types during learning also prepares students for varied assessment formats they might encounter in exams or real-life scenarios.

**** Task 4. Similar questions generation***

Generating similar questions is another beneficial task. ChatGPT can create variations of a question while retaining the core concept (see Figure 10). These similar questions can be valuable for reinforcing understanding and providing additional practice to students (see Figure 11).

TR

Lengths of Fish (in inches)

8 9 9 9 10 10 11

11 12 12 12 12 13 13

13 14 14 15 15 16 24

The table above lists the lengths, to the nearest inch, of a random sample of 21 brown bullhead fish. The outlier measurement of 24 inches is an error. Of the mean, median, and range of the values listed, which will change the most if the 24-inch measurement is removed from the data?

- A) Mean
- B) Median
- C) Range
- D) They will all change by the same amount.

The above question is extracted from a SAT Practice Test. Create a similar question.

Figure 10. Original question

Question:

Weights of Gym Equipment (in kilograms)

 Copy code

4.54 5.44 6.80 9.07 9.98 11.34 12.70

5.44 5.90 6.80 8.62 9.53 10.43 12.25

6.35 6.80 7.71 8.62 9.98 10.89 14.51

The table above represents the weights, in kilograms, of different gym equipment. One weight measurement of 14.51 kilograms was incorrectly recorded; it should have been 10.43 kilograms. If this error is corrected, which measure of central tendency will change the most?

- A) Mean
- B) Median
- C) Mode
- D) They will all change by the same amount.

Figure 11. Similar question generated

While the questions use different datasets (fish lengths and gym equipment weights), they employ a comparable structure. In each case, there's an outlier or incorrectly recorded value that impacts the statistical measures. They aim to assess the test-taker's understanding of how outliers or inaccuracies impact measures of central tendency.

2.3.3. Sample test

This section details the process undertaken, including three main steps, to create a customized quiz paper tailored to the educational context of Vietnamese students.

*** Step 1: Make a request**

- Identifying Learning Objectives: The first step involves a comprehensive analysis of the learning objectives and course content related to Statistics.

- Establishing Cognitive Levels (see Figure 12): To ensure a diverse range of questions, four levels of cognitive learning are defined: easy (memorizing concepts), medium (calculating summary statistics), hard (explaining real-life implications), and challenging (concluding). These levels aim to gauge students' knowledge and application skills progressively.



You

There will be four levels of cognitive learning in my test:

- Level 1 (easy): Remembering the concepts of summary statistics for measuring central tendency of ungrouped data: mean, mode, median, quartiles;
- Level 2 (medium): Calculating the summary statistics for measuring the central tendency of ungrouped data: mean/average, mode, median, quartiles.
- Level 3 (hard): Explaining the meaning and role of the above summary statistics of the sample data in real life and Selecting the appropriate summary statistic for measuring the central tendency of some ungrouped data;
- Level 4 (challenging): Indicating the conclusions based on the meaning of the above summary statistics of the sample data.

Figure 12. Establishing Cognitive Levels

- Question Type: Multiple-choice questions are selected as they provide a structured format while allowing for diversity in content and difficulty.

- Integration of Vietnamese Context: Recognizing the importance of cultural relevance, the quiz incorporates a passage stimulus for questions from Level 2 to Level 4, directly related to the Vietnamese educational and social context. Data tables are included in questions 5 and 8.

* *Step 2: Iterative Refinement and Evaluation*

The quiz paper undergoes iterative refinement, involving continuous evaluation and feedback. The initial version is scrutinized against certain criteria, including clarity, relevance, challenge level, and effective use of reference material. The iterative refinement ensures the final quiz paper is of the highest standard.

* *Step 3: Ask for Latex code*

The final step is to format the quiz using LaTeX, which helps ensure that the quiz looks neat and professional (see Figure 13). It ensures the quiz fits on two pages as per the user's request. Each question, answer choice, and table were meticulously formatted to enhance readability.



You

provide me Latex code of the whole test above with format fitting 2 pages



ChatGPT

Certainly! Below is the LaTeX code for the entire test paper with formatting designed to fit on two pages. You can use this code in a LaTeX editor or document to create your quiz paper:

```
latex Copy code

\documentclass[letterpaper,10pt]{article}
\usepackage[utf8]{inputenc}
\usepackage[margin=1in]{geometry}
\usepackage{enumitem}
\usepackage{multirow}
\usepackage[table]{xcolor}

\title{Statistics Quiz}
```

Figure 13. Ask for the Latex code

The whole test is shown in Figure 14. Creating mini-tests with ChatGPT reinforces the effectiveness of the proposed procedure while also saving educators time in the test development process.

Statistics Quiz

Instructions:

- This test contains 10 questions.
- You have 20 minutes to complete the test.
- Answer all questions.
- For multiple-choice questions, select the letter corresponding to your chosen answer.
- Good luck!

Level 1 (easy - Remembering Concepts):

1. In a recent mathematics class, the teacher discussed measures of central tendency, including the concept of the "middle" value in a dataset. Which of the following measures of central tendency represents the "middle" value in a dataset?
(a) Mean (c) Median
(b) Mode (d) Quartiles
2. During a statistics lesson, the teacher explained that the mode is the value that appears most frequently in a dataset. What is the definition of the mode in a dataset?
(a) The sum of all values divided by the number of values. (c) The middle value in the dataset.
(b) The value that appears most frequently. (d) The range between the minimum and maximum values.
3. In a recent test, students were asked to calculate various summary statistics for a given set of data, including quartiles. Which of the following is a measure of central tendency that divides the data into four equal parts, with each part containing 25% of the data?
(a) Mean (c) Median
(b) Mode (d) Quartiles

Level 2 (medium - Calculations):

4. In a recent mathematics competition held in Hanoi, students were asked to solve a series of problems. The scores of ten participants are as follows:

Student	A	B	C	D	E	F	G	H	I
Score	85	92	78	88	90	86	89	91	84

What is the median score for the students in the mathematics competition?

- (a) 88 (c) 88.5
(b) 89 (d) 87.5

5. A class of 6th-grade students in Da Nang recently took a science quiz. Their scores are as follows: 85, 90, 88, 92, 78, 89, 91, 86, 84, 87. Calculate the mean score of the 6th-grade students on the science quiz.

- (a) 86 (c) 88.5
(b) 87 (d) 88

6. In a survey conducted among primary school students in Ho Chi Minh City, students were asked to choose their favorite colors from a list. The responses are as follows: Red, Blue, Green, Green, Yellow, Blue, Red, Green, Yellow, Blue. What is the mode of the students' favorite colors?

- (a) Red (c) Green
(b) Blue (d) Yellow

7. A recent study in Hue City examined the monthly incomes of workers in a factory. The data is shown below:

Worker	Monthly Income (in VND)
A	5,000,000
B	4,500,000
C	4,800,000
D	5,200,000
E	4,700,000
F	5,100,000
G	4,600,000
H	5,300,000
I	4,900,000
J	5,000,000

Calculate the first quartile (Q1) of the monthly incomes of the workers.

- (a) 4,500,000 VND (c) 5,000,000 VND
(b) 4,800,000 VND (d) 5,100,000 VND

Level 3 (hard - Real-Life Context):

8. A recent survey on student commute times to school in Hanoi, Vietnam, found that the median commute time was 35 minutes. What does the median commute time of 35 minutes mean in the context of student commutes?

- (a) It represents the average commute time for students. (c) It indicates that half of the students have commute times shorter than 35 minutes.
(b) It is the longest commute time recorded. (d) It has no relevance to student commutes.

9. An agricultural study in rural Vietnam recorded the mode of crop yields for a specific region. How can understanding the mode of crop yields assist local farmers and policymakers in making informed decisions?

- (a) It helps in calculating the average crop yield.
(b) It identifies the crop with the highest yield.
(c) It provides insight into the most common crop yield in the region.
(d) It has no relevance to agricultural decisions.

Level 4 (challenging - Drawing Conclusions):

10. A study on the heights of students in a Vietnamese high school found that the mean height was 165 cm. Based on this information, what conclusions can you draw about the typical height of students in this school? Explain your answer.

- (a) The typical height is exactly 165 cm for all students. (c) The typical height is close to 165 cm on average.
(b) The majority of students are shorter than 165 cm. (d) The study is inconclusive regarding student heights.

3. Conclusions

This paper studies the requirements of Grade 10 Statistics knowledge, particularly focusing on measures of central tendency for ungrouped data and aligning these requirements with Bloom's Taxonomy levels range from Understanding to Applying and Analyzing.

This study emphasizes the importance of evaluating students' grasp of statistical knowledge within practical contexts. The paper introduces a structured two-step procedure utilizing ChatGPT 3.5 to design real-world Statistics problems, complemented by four additional tasks aimed at enhancing question quality. Moreover, a comprehensive three-step approach is proposed to create a cohesive mini-test. Some limitations of ChatGPT in generating questions include a lack of diversity in contexts, the potential for a shortage or redundancy of questions, and the generation of questions that may not relate to statistics.

Looking ahead, further analysis opportunities emerge, including extending the process of question generation across diverse branches, implementing and assessing these questions in real educational settings, and conducting a comparative analysis between ChatGPT 3.5 and its successor, ChatGPT 4. These analyses aim to enrich our understanding of AI's evolving role in educational assessments. Additionally, the study paves the way for critical inquiries, such as how mathematics educators can provide feedback to AI systems like ChatGPT for improved question generation quality. It also provokes consideration of the potential long-term impact of integrating AI-generated questions on sustained improvements in student learning outcomes.

Acknowledgment: Le Thuy Dieu was funded by the Master, PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), code VINIF.2023.ThS.023

REFERENCES

- [1] Van den Heuvel-Panhuizen M, (2005). The role of contexts in assessment problems in mathematics. *For the Learning of Mathematics*, 2(2), 2-9.
- [2] Christine F, Gary K, Denise M, Jerry M, Roxy P, Mike P & Richard S, (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report*.
- [3] Barab S, Zuiker S, Warren S, Hickey D, Ingram-Goble A, Kwon E & Herring C, (2007). Situationally embodied curriculum: Relating formalisms and contexts, *Science Education*, 91, 750-782, <https://doi.org/10.1002/sce.20217>.
- [4] Ministry of Education and Training, (2018). Mathematics General Education Program. Circular No. 32/2018/TT-BGDĐT, Hanoi.
- [5] David N, Michelle H & Michelle N, (2013). Using real-life data when teaching Statistics: Student perceptions of this strategy in an introductory Statistics course. *Statistics Education Research Journal*, 12(2), 59-70.
- [6] Andriani L & Fauzan A, (2019). The impact of RME-based design instructional on students' mathematical communication ability. *International Journal of Scientific & Technology Research*, 8(12), 2646-2649.
- [7] Gil E & Ben-Zvi D, (2011). Explanations and context in the emergence of students' informal inferential reasoning. *Mathematical Thinking and Learning*, 13(1-2), 87-108.
- [8] Das B, Majumder M, Phadikar S & Sekh A, (2021). Automatic question generation and answer assessment: A survey. *Research and Practice in Technology Enhanced Learning*, 16(1), 1-15. <https://doi.org/10.1186/s41039-021-00151-1>.

- [9] Floridi L & Chiriatti M, (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30, 681-694. <https://doi.org/10.1007/s11023-020-09548-1>.
- [10] Ministry of Education and Training, (2018). General Education Program - Comprehensive Program (Issued together with Circular No. 32/2018/TT-BGDDT on December 26, 2018).
- [11] Hernandez-Martinez P & Vos P, (2018). Why do I have to learn this? - A case study on students' experiences of the relevance of mathematical modeling activities. *ZDM*, 50, 245-257.
- [12] Pauline V, (2018). How Real People Really Need Mathematics in the Real World - Authenticity in Mathematics Education. *Education Sciences*, 18(4), 195, <https://doi.org/10.3390/educsci8040195>.
- [13] Palm T, (2006). Word problems as simulations or real-world situations: A proposed framework. *For the Learning of Mathematics*, 25(1), 42-47.
- [14] Ppbi, Umt - Pusat Pembangunan Bakat Dan Inovasi, (2023). ChatGPT in student assessment. *Youtube*, https://www.youtube.com/watch?v=AoIpznWzSPQ&t=2792s&ab_channel=PPBI%2CUMT-PUSATPEMBANGUNANBAKATDANINOVASI.