ChatGPT as a Pedagogical Tool for Teaching Statistical Distributions: An Empirical Classroom Study

David Tian[1]* iD, Jinju Lee[2]

Abstract

This study was conducted at an international high school in South Korea to evaluate the effectiveness of incorporating ChatGPT as a tool for teaching data distributions. Using a pretest/posttest design, students were given a brief, traditional lecture on various statistical distributions followed by a diagnostic test. Subsequently, they engaged with ChatGPT to generate, and graph simulated datasets. The investigation involved a small sample of seven students over a 100-minute class, concluding with a summative assessment. Results indicate substantial improvement in most students, with a p-value of approximately 0.049, demonstrating the potential of AI-assisted learning in the classroom.

Keywords: ChatGPT, Education Technology, AI-assisted learning, data science pedagogy, Generative AI.

1. Introduction

The public release of ChatGPT has spurred a flurry of interest among educators, many of whom are concerned about students using generative AI to engage in academic dishonesty or otherwise undermining the educational process [1]. Additionally, in a review article, Gupta et al. (2023) have found that a common concern among scholars is that ChatGPT frequently produces factual and logical errors, which is counterproductive to the goal of providing accurate information in education [2]. However, as with many technological innovations throughout the years, the advent of ChatGPT also brings with it tremendous potential to augment the classroom experience for students. Indeed, some authors, including Ellis and Slade (2023), have already begun investigating the use of ChatGPT as a tool to enhance the teaching of statistics and data science [3].

In line with this emerging trend, the present study explores the use of ChatGPT as an interactive, hands-on, and expedient learning tool in a high school setting, with a focus on teaching statistical distributions. The study evaluates the effectiveness of using ChatGPT as a classroom supplement, leveraging the use of AI to bolster learners' understanding of statistical concepts and increase their analytical thinking skills. By analyzing the impact of these AI-assisted teaching methods, this study aims to contribute to the discourse surrounding the role of technology in reshaping educational paradigms for the new digital age.

The article is organized as follows: first, there will be a discussion of the methodology employed, including the initial assessment, the interactive learning session, and the summative assessment. Then, the results of both assessments will be summarized, accompanied with a discussion about them. Following the discussion of the results, there will be a quantitative analysis using statistical methods to ascertain the effectiveness of the intervention. Finally, the present study will conclude with implications about its generalizability and suggestions for follow-up studies.

2. Methodology

This study employed a pretest/posttest design to evaluate the efficacy of ChatGPT as a supplemental tool in teaching statistical distributions to high school students. The methodology was structured in three primary phases: initial assessment, interactive learning session, and final assessment, conducted over two fifty-minute sessions held back-to-back. Participants were already enrolled in the class AP Computer Science Principles, which is the context in which this study was implemented. This non-randomized selection is recognized as an inherent limitation of the study. Nevertheless, to prevent bias, all students received the same materials, and both assessments had objective answers, providing a consistent intervention and minimizing the possibility of instructor bias.

2.1 Initial Assessment

The study began with a diagnostic test, given after a traditional lecture-style lesson, to assess the students' baseline understanding of statistical distributions. Consistent with previous research that underscores the importance of using culturally relevant data when teaching statistics,[4] the test included questions involving situations that would be of interest to students at an international high school. For example, three of the questions, with the correct answers given in square brackets, were:

1) What kind of distribution would you expect for the scores of an exam at an international high school where half the students studied, and half the students did not? [bimodal distribution]

2) What kind of distribution would you expect for scores on an easy quiz on which the vast majority of students get perfect or near-perfect scores, and a few get lower scores? [left-skewed distribution]

3) What kind of distribution would you expect for the heights of students at an ordinary high school? [normal distribution]

The diagnostic test served to gauge students' initial competency before the use of ChatGPT.

2.2 Interactive Learning Session

Following the initial assessment, students were then instructed to use ChatGPT. In this phase, students were instructed to give prompts to ChatGPT along the lines of:

Create Python code that generates and graphs a realistic dataset for the test scores of a final exam in a large introductory college course with 200 students.

[1] School of Engineering and Applied Science, University of Pennsylvania, USA
[2] Daegu University, Gyeongsan, South Korea
* Corresponding Author: David Tian
E-mail address: dktian@seas.upenn.edu, ORCID: https://orcid.org/0000-0002-3456-3691

These datasets were then graphed in an integrated development environment, allowing students to visualize the shapes of the distributions directly. ChatGPT was further used to obtain explanations about the characteristics and implications of these distributions, along with clarifications of why a certain situation produced a particular distribution of data, which facilitated a deeper understanding.

2.3 Final Assessment

During the final twenty minutes of class, students were given a summative assessment. It was structured the same way as the initial test with similar questions, the results of which provided data to compare against the pre-test results, enabling an evaluation of the effectiveness of using ChatGPT in teaching statistical concepts.

3. Results

The results of the assessments are presented in Table 1. Out of the seven students, six of them received scores that were higher on the final assessment than on the initial assessment, and in some cases, significantly so. One student, however, namely Student 3, received a score that was much lower on the summative assessment than the diagnostic. The scores for both tests had a maximum possible value of 20. The mean for the diagnostic was approximately 7.43, and the mean for the summative assessment was 14 - a nearly twofold improvement. A visualization of the data is presented in Figure 1.

Table 1. Comparative Table of Diagnostic and Summative Assessment Scores

Student	Diagnostic Score	Summative Score
1	7	17
2	5	9
3	12	6
4	7	17
5	5	16
6	3	18
7	13	15
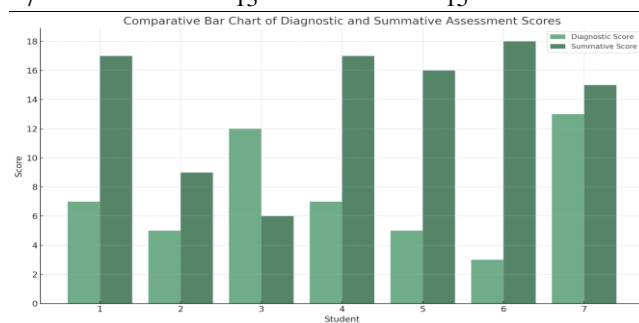
Figure 1. Bar Chart of Outcomes

4. Analysis

The data were analyzed using a paired sample t-test to determine the significance of score changes from the diagnostic to the summative assessment. In doing so, two assumptions were made: 1) that the data are normally distributed and 2) that each participant was independent. With six out of seven students showing improvement and a calculated p-value of approximately 0.049 - statistically significant at the 0.05 alpha level - the results suggest that the study's intervention had a net positive effect on student outcomes. However, the precipitous decline in Student 3's performance is notable. Without a clear external factor to account for this deviation, one can only speculate about various internal factors such as fatigue at the end of two class periods. Anecdotal observations suggest that Student 3's performance drop may stem from feelings of discouragement after seeing preliminary results, which, unrelated to the use of ChatGPT, could have lessened subsequent effort in the later assessment.

Without student 3's anomalous data, the p-value drops to under 0.0069, providing strong evidence supporting this study's hypothesis. Despite this aberration, the overall trend supports the theory that AI-assisted instruction can enhance learning, indicating that ChatGPT, when integrated with a traditional teaching approach, can reinforce and deepen students' understanding of statistical distributions and potentially other subjects as well. In any case, the results of this study demonstrate the potential for AI to serve as a valuable addition in educational settings, offering interactive and personalized learning experiences that can complement conventional pedagogical methods.

5. Conclusions

Empirical evidence gathered from this study indicates that the careful integration of ChatGPT into the classroom holds promise as an effective teaching strategy. In particular, given the current concerns about factual accuracy, method-oriented classes (e.g. programming or statistics) could benefit more greatly than content-oriented classes (e.g. history or geography). While the study is limited by its small sample size and one anomaly, the overall data portray a positive trajectory towards enhanced educational outcomes through the utilization of AI. Additionally, as pretest/posttest research designs have been met with methodological concerns in the past,[5] further research with a larger sample size, experimental methods, and randomization can help confirm these findings and explore other facets of AI-assisted learning. Nonetheless, given the compelling results of this research, there is a clear impetus for the inclusion of generative AI in a teacher's pedagogical toolkit. As the educational landscape evolves, embracing and optimizing innovative tools like ChatGPT may be key to educating the digital generation effectively and efficiently.

Conflict of Interest Statement

The authors declare no conflict of interest.

References

[1] Debby R. E. Cotton, Peter A. Cotton & J. Reuben Shipway, Chatting and cheating: Ensuring academic integrity in the era of ChatGPT, *Innovations in Education and Teaching International*, 61:2 (2024) 228–239,

[2] Bulbul Gupta, Tabish Mufti, Shahab Saquib Sohail & Dag Øivind Madsen ChatGPT: A brief narrative review, Cogent Business & Management, 10:3 (2023).

[3] Amanda R. Ellis & Emily Slade, A New Era of Learning: Considerations for ChatGPT as a Tool to Enhance Statistics and Data Science Education, *Journal of Statistics and Data Science Education*, 31:2 (2023) 128–133.

[4] Travis Weiland & Immanuel Williams (2023) Culturally Relevant Data in Teaching Statistics and Data Science Courses, *Journal of Statistics and Data Science Education*,

[5] Emma Marsden & Carole J. TorgersonSingle group, pre- and post-test research designs: Some methodological concerns, *Oxford Review of Education*, 38:5 (2012) 583–616.