Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Role-based Prompting Technique in Generative AI-Assisted Learning: A Student-Centered Quasi-Experimental Study

Houda Louatouate[1]✉ and Mohammed Zeriouh[2]
[12]Department of Education Sciences and Technologies of Teaching, the Higher School of Teachers (ENS), Moulay Ismail University, Morocco
Corresponding Author: Houda Louatouate, **E-mail**: h.louatouate@edu.umi.ac.ma

| ABSTRACT

The education landscape has known remarkable transformations with the emergence of AI which has shown great potential in optimizing different educational processes provided its solutions are utilized effectively. Hence, it has become a necessity to open eyes to practices promoting worldwide efficient use. To this aim, the present paper investigates role-based prompting, its importance in enhancing the output quality of models in AI-assisted learning, and students' satisfaction with the overall performance with and without using role-based prompts. To achieve the required results, the present study adopts a quasi-experimental pretest-posttest research design with a student-centered approach. The sample of this study includes N=43 education bachelor's students of the Higher School of Teachers – Moulay Ismail University, whose ratings were measured before and after the researchers' intervention. For data analysis, following the ordinal non-normally distributed nature of data, the study adopts Wilcoxon Signed-Rank, a non-parametric test for paired samples conducted using both SPSS 25 and Python code executed on Google Colab coding space for robust statistical transparency and evidence. Results revealed a statistically significant difference in output quality before and after using role prompting technique. Additionally, results demonstrated that role-based prompts optimize the output quality in terms of clarity, depth, professionalism, insightfulness, innovativeness, relevance, and generosity. It was also found that students' satisfaction with the output quality significantly increases with the use of role-based prompts. Furthermore, the paper at hand sheds light on limitations and recommendations to guide future research projects in the field or in fields that relate to it.

| KEYWORDS

AI-assisted learning, models, output quality, role-based prompting, satisfaction

| ARTICLE INFORMATION

ACCEPTED: 12 March 2025 **PUBLISHED:** 17 April 2025 **DOI:** 10.32996/jcsts.2025.7.2.12

1. Introduction

The growing importance of artificial intelligence in today's world is unquestionable as the adoption of AI, similar to typical technological tools previously, is slowly becoming a necessity rather than a luxury. Likewise, the field of education has witnessed a constantly evolving interest in discovering and examining limitless generative AI features, tools, and their applications, as well as the ways in which they can contribute effectively to the development of the educational sector in general. Unarguably, the new wave of AI is accompanied by multiple questions on the exact skills that can be developed and adapted to successfully interact with these intelligent systems and harness their capabilities. One way is the use of effective prompting techniques adequate to refine chatbot models' performance. It is noteworthy that good prompting skills represent the stepping-stone towards predicting how well a chatbot model will react to well-rounded prompts, especially textual ones. However, we need to recognize how challenging it can be to develop these specific skills due to the novelty of prompting as an aspect of human-machine interaction mechanisms.

1.1. Research questions

The present study aims to answer two main questions:

1- To what extent does the use of role-based AI prompts improve AI models' output quality in AI-assisted learning from students' standpoint?
2- Does students' satisfaction with the quality of AI models' output increase with the use of role-based AI prompting technique?

1.2. Hypotheses

- **Output Quality Hypotheses**

H_0: Using role-based prompts in input content does not significantly affect the quality of the output.

H_1: Using role-based prompts in input content significantly affects the quality of the output.

- **Satisfaction Hypotheses**

H_0: There is no significant difference between students' satisfaction with output quality before and after using role-based prompts.

H_1: There is a significant difference between students' satisfaction with output quality before and after using role-based prompts.

1.3. Research objective

Developing students' skills for a refined use of generative AI has become a central issue in the literature. This study aims to shed light on the importance of prompting in optimizing the quality of responses produced by GenAI models to open eyes to the game-changing role possessing prompting skills plays in reshaping student's AI-enhanced learning experience.

2. Review of the literature

2.1. Generative AI (GenAI)

Generative AI is an aspect of Artificial Intelligence covering a wide range of systems that rely on particular developed technology fields of study including Machine Learning ML which refers to a set of Algorithms capable of self-development based on existing data in order to produce content characterized by its partial uniqueness as it recycles what it learns over time to reproduce potentially described valuable content. In this concern, Feuerriegel et al. (2024) state, "The term generative AI refers to computational techniques that are capable of generating seemingly new, meaningful content such as text, images, or audio from training data." (p. 111). Feuerriegel et al. (2024) provide a concise explanation of generative AI which is mainly concerned with content they described as "seemingly new". The choice of words here is extremely meaningful as generative AI-produced responses can be perceived as unique and original when they are generated based on already existing data.

In essence, generative AI's ability to process information and produce textual content using natural language makes students in the context of education become rapidly familiar with its use. In the same regard, Feuerriegel et al. (2024) argue, "ChatGPT's ease of use also for non-expert users was a core contributing factor to its explosive worldwide adoption." (p. 115). In other words, one very important factor behind ChatGPT's wide use as a popular aspect of generative AI lies in its smooth nature allowing users with little expertise in AI and ITs to equally have access to it. In the same regard, Michel-Villarreal et al. (2023) mention, "Given the novelty of ChatGPT, existing literature about its use within higher education is limited and largely hypothetical or speculative." (p.14). Michel-Villarreal et al. (2023) highlight the lack of scholastic guidelines in the literature shaping the use of ChatGPT in higher education which can operate as a central issue in generative AI's use in higher education as students tend to have very limited knowledge and skills in dealing with generative AI models which can raise misuse issues.

Mittal et al. (2023) clarify, "Specific GenAI models, such as GANs, are known to be unstable, making it challenging to control their behavior. They often fail to produce the expected outcomes, and understanding the reasons behind these deviations can be complex and difficult to detect." (p.16). Mittal et al. (2023) tackle a very important matter facing students as well in generative AI related to the complexity of chatbot models' behavior which renders performance predictions -in many cases- a mission impossible. However, as users cannot control a GenAI model's behavior, there are discussions in the literature tackling the control over user's input. Regarding input content, Oppenlaender et al. (2024) mention, "This encompasses not only the basic knowledge of the relevant language syntax but also the strategic use of prompt modifiers—elements that refine or alter the direction of generative outputs." (p. 5). In their statement, we can grasp the idea that crafting input content using appropriate language syntax is not enough as other element can interfere with produced content's direction.

2.2. Prompt engineering and its importance in education

Prompting refers to the act of crafting instructions for a model to make it generate a desired output. In a study conducted on Generative AI and Prompt Engineering, Bozkurt and Sharma (2023) mention, "Prompt engineering refers to the process of designing, crafting, and refining contextually appropriate inputs or questions in order to elicit specific types of responses or behaviors from an AI language model." (p. 2). Bozkurt and Sharma (2023) provide a very clear definition describing prompt engineering which is not limited to specific fields but everyone can act as a prompt engineer while trying to form the most efficient input prompts. In the same regard, Bozkurt and Sharma (2023) state, "The key to facilitating effective communication and interaction between humans and generative AI lies in skillfully crafting prompts." (p.4). In an attempt to provide a definition for prompt engineering skill as a novel type of skills in a completely new discipline, Oppenlaender et al. (2024) on their turn claim, "we define 'skill in prompt engineering' as the ability to effectively utilize language and prior knowledge to craft prompts that guide generative models towards desired outputs." (p. 5).

We should admit that the use of generative AI in general and in education, in particular, has given birth to new skills' criteria which can shape students' interactions with GenAI models. Bozkurt and Sharma (2023) claim, "The goal of prompt engineering is to optimize how the model responds based on the structure, content, and tone of the question, thereby facilitating more accurate, useful, or engaging interactions." (p. 2). The goal behind mastering prompt engineering was put in simple words by Bozkurt and Sharma (2023) who clarify that chatbots' performance can possibly be optimized through well-rounded prompts. In this matter, higher education students in many cases do not discern the existence of prompt engineering as a structured discipline and do not possess fundamental basic knowledge about the possibility of having control over the quality of a chatbot's performance and the relevance of the content it produces primarily tied with their input quality as users, thinking that generated content quality and relevance criteria are mainly associated with the system's programming not their own skills.

In essence, chatbot models are designed to imitate human intelligence. Consequently, they produce content which relies heavily on how user's request was expressed and perceived just as an extremely intelligent human being would do. Many researchers in the literature to hold a positive attitude towards the human-models interaction with the use of efficient prompts, Park and Choo (2024) mention, "While prompt engineering has great potential to improve human interactions with generative AI tools and to optimize the outputs of AI models, few resources are available, or even relevant, to special educators" (p. 5).

In addition to what has been said, Park and Choo (2024) shed light on an important issue related to the lack of research studies aiming at refining educators' skills in prompting. However, students are also in need of special guidelines to enhance their prompting skills. In a study conducted by Theophilou et al. (2023), they mention, "When comparing the two ChatGPT activities, the initial one without prompting strategies and the second one after being introduced to prompting, the students rated the second interaction with ChatGPT clearer, more natural and better than the initial attempt." (p.10). We can clearly grasp, based on this statement, the impact the use of prompting strategies can generally have on the human-machine interaction experience especially in education. Among prompting strategies, we find the role-based technique which consists of assigning the AI-model a specific role. Chatbot models indeed act as the assigned role and provide responses depending on it. For instance, a chatbot model can be asked to act as an expert in education. Subsequently they'll act accordingly.

Unfortunately, and due to the novelty of the subject being discussed, very few studies have dealt with some prompting strategies in a detailed manner to decide whether they are indeed effective or not especially for students who are in extreme need of similar skills' development. In the same context, Michel-Villarreal et al. (2023) state, "A key takeaway point is the urgent need for empirical research that delves into best practices and strategies for maximizing the benefits of GenAI, as well as user experiences, to understand students' and academics' perceptions, concerns, and interactions with ChatGPT" (p.14). The gap associated with GenAI prompting and required skills in the literature is clearer than any other one. There is a great need for further studies aiming at tackling these matters from different angles to come-up with existing difficulties and solutions to them.

3. Methodology

3.1. Research design

In this study, it was important to measure the selected model's (GPT-4) output quality in generative AI-Assisted learning from students' standpoint with and without using role-based technique, which is well-known for its efficacy in raising chatbot systems' potential at different levels. Thus, involving students in similar investigations will help researchers identify key elements directly connected with students' learning and understanding. To this aim, the present paper adopts a quasi-experimental one-group pretest-posttest research design to assess the quality of ChatGPT's generated content without and with one prompting technique based on a student-centered approach. We relied on students' evaluations which is a common approach to measure the quality of information and the quality of systems. Data was analyzed using Wilcoxon Signed-Rank test for two related samples using SPSS version 25 and Python code.

3.2. Sample

The sample of this study includes N= 43 bachelor's students of the Higher School of Teachers – Moulay Ismail University, who were selected to participate in this study following a randomized selection procedure. The same group took both the pre-test and post-test to minimize inaccuracy due to differences in levels of understanding and background knowledge between two different groups which can interfere with the quality of the obtained results.

3.3. Research Instrument

The study is built on a pretest-posttest design. The same form was administered before and after researchers' intervention representing both the pre-test and post-test material in order to accurately measure the same constructs before and after the intervention**.** The test form consisted of 8 questions, seven of which represented the first part of the test and aimed at measuring ChatGPT's output (generated content) quality based on seven main elements: Clarity, depth, professionalism, insightfulness, innovativeness, relevance, and generosity. The second part which was also the last one aimed at measuring students' overall satisfaction. All questions were measured on 4-point Likert scale to help participants provide well-structured ratings.

3.4. Procedure

Participants were first asked earlier to kindly bring their laptops as one of the experiment requirements. They were also informed that their participation was completely voluntary and that they had the complete right to withdraw at any stage of the data collection process. To ensure highly accurate responses, participants were reassured that there are no right or wrong answers and that they should provide their mere independent data without relying on anyone else's perspectives. At the beginning students were also informed that we had to craft a request together. We agreed on "Tell me how to start a business" which was going to be their user's input and an initial request that we were going to have ChatGPT generate a response for. Students were asked to take their time to read carefully the generated content and answer the pre-test questions afterwards. Once they finished, participants were asked to copy the used request and open a new ChatGPT conversation again, paste the same exact request but this time they were asked to add "You are an expert in X" at the beginning of the initial input. X can vary depending on the corresponding domain be it business, research, teaching, cinematography etc. Students had to modify the initial prompt and use it as a new input. In our case, the second version of the prompt using role-based technique was as follows: "You are an expert in entrepreneurship. Can you tell me how to start a business". Students were asked to take their time to carefully read the newly generated content using the refined version of the prompt. Post-test was administered and similarly to the pre-test procedure, students were asked to rate the second output based on 7 main criteria and rate their overall satisfaction at the end. The type of prompt crafted was a random one used to test the importance of role-based prompts in general which is applicable to their specialty as well but does not directly relate to their specific field of study or personal life. Data was gathered and results were analyzed.

3.5. Data analysis procedure

Gathered data was analyzed using SPSS 25. Following the ordinal nature of the gathered non-normally distributed data of this study, the paired Wilcoxon Signed-Rank Test was conducted as a non-parametric test on SPSS 25 too which is universally known for its robust and unbeatable statistical performance. However, SPSS usually provides values with no more than three decimal places and due to the extremely small p value we got at different points and which already suggests the highly significant existing difference, we decided to conduct Wilcoxon Signed-Rank Test using Python Code, which was executed on Google Colab Online coding space and which has provided exact p values for more statistical accuracy.

4. Results

To answer the two questions raised at the beginning of this paper, this section is divided into two major parts: Output quality and satisfaction. Each part aims at presenting results adequate to answer the appropriate research question.

4.1 Output quality

4.1.1. Clarity

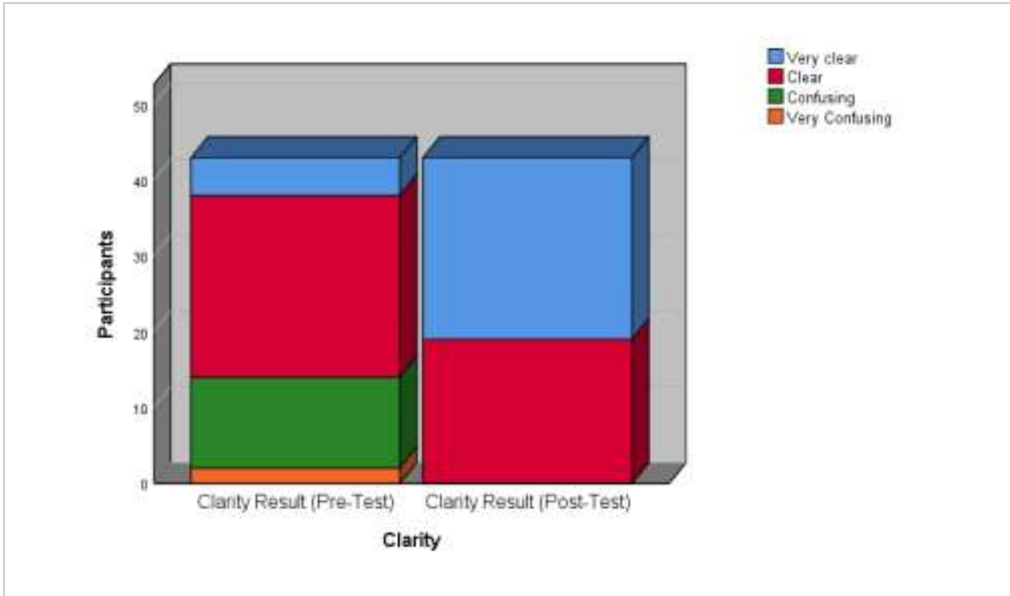

Figure 1. *Results of output clarity before and after using role prompting*

As shown in Figure 1 which presents ChatGPT's generated response clarity results before and after using role-based AI prompting technique. Results of the pre-test show that 55.8% found the output to be clear, 32.6% claimed it to be confusing and very confusing, whereas only 4.7% used very clear option to rate the quality of the response they got before using role-based prompt. On the other hand, results of post-test which aimed at measuring the response clarity after using role-based technique are distinguished by the disappearance of 'Confusing' and 'very confusing' claims as a majority of 55.8% described the output to be very clear (compared to 4.7% in the pre-test), followed by 44.2% of students who reported the output to be clear, representing 100% of students, confirming the rise in overall quality output in general and clarity in particular.

To statistically test the significance of differences in clarity, we performed the Wilcoxon Signed Rank Test under SPSS version 25. We got the following results: $p < 0.001$, $|z| = 5.31$. Results show a strong significance according to SPSS statistics which provides numbers with only three decimal places. For although p value in this case shows a strongly large significance as it is extremely low $p < 0.0001$ ($p = 0.000$ according to SPSS statistics). For stronger statistical transparency, we decided to run the paired Wilcoxon Signed-Rank Test using Python and Google Colab coding environment to extract the exact small value of p. Results were as follows:

$$p = 2.31743731097297e\text{-}07$$

$$p = 0.0000002317$$

```
import numpy as np
from scipy.stats import wilcoxon
# data (before and after values)
before = [2, 2, 3, 1, 2, 1, 2, 2, 2, 2, 3, 2, 1, 2, 1, 2, 2, 2, 1, 3, 0, 2, 2, 2, 1, 2, 2, 3, 1, 2, 1, 2, 1, 2, 2, 2, 3, 1, 2, 1, 2, 3, 1]
after = [3, 3, 3, 3, 3, 2, 3, 3, 3, 3, 3, 2, 2, 3, 2, 3, 3, 3, 2, 3, 2, 3, 3, 2, 3, 3, 3, 2, 3, 3, 3, 3, 3, 2, 2, 3, 3, 3, 3, 2, 3, 3, 2]
# Perform Wilcoxon signed-rank test
statistic, p_value = wilcoxon(before, after)
# Output the p-value
print("P-value:", p_value)
```

```
P-value: 2.31743731097297e-07
```

Figure 2. *Python code for the generation of the exact p value.*

The matter was settled as shown in Figure 2. We achieved once again strong statistical evidence suggesting the existence of a difference between GPT's output clarity before and after using role-based prompting. As it is crucially essential to illustrate the direction of the obtained statistical results. The paired Wilcoxon Signed-Rank Test was run again under SPSS 25 to display the table of ranks showing the exact direction of the difference in the quality of output clarity before and after using the AI-prompting technique. The table below shows the distribution of positive ranks, negative ranks, and ties (participants who stuck to the same rating and have not shown any perceived difference between pre-test and post-test).

Table 1. *Wilcoxon Signed-Rank Test Table of Ranks (Output Clarity)*

		N	Mean Rank	Sum of Ranks
ClarityPosttest - ClarityPretest	Negative Ranks	1	16,00	16,00
	Positive Ranks	33	17,55	579,00
	Ties	9		
	Total	**43**		

As shown in Table 1, out of 43 participants, results show that there are 33 Positive Ranks suggesting that 33 of participants have reported an increase in output clarity after using role-based technique compared to the initial output, 1 negative Rank suggesting that 1 participant has reported a decrease in in output clarity, and 9 participants out of 43 who kept the same position before and after using role-based prompting without reporting any change any clarity after researchers' intervention.

Wilcoxon Signed-Rank Test suggests a highly significant difference between pre-test and post-test score. We conclude that results suggest a very strong rejection of the null hypothesis for the output quality increase before and after using role-based prompting technique for clarity.

4.1.2. Depth

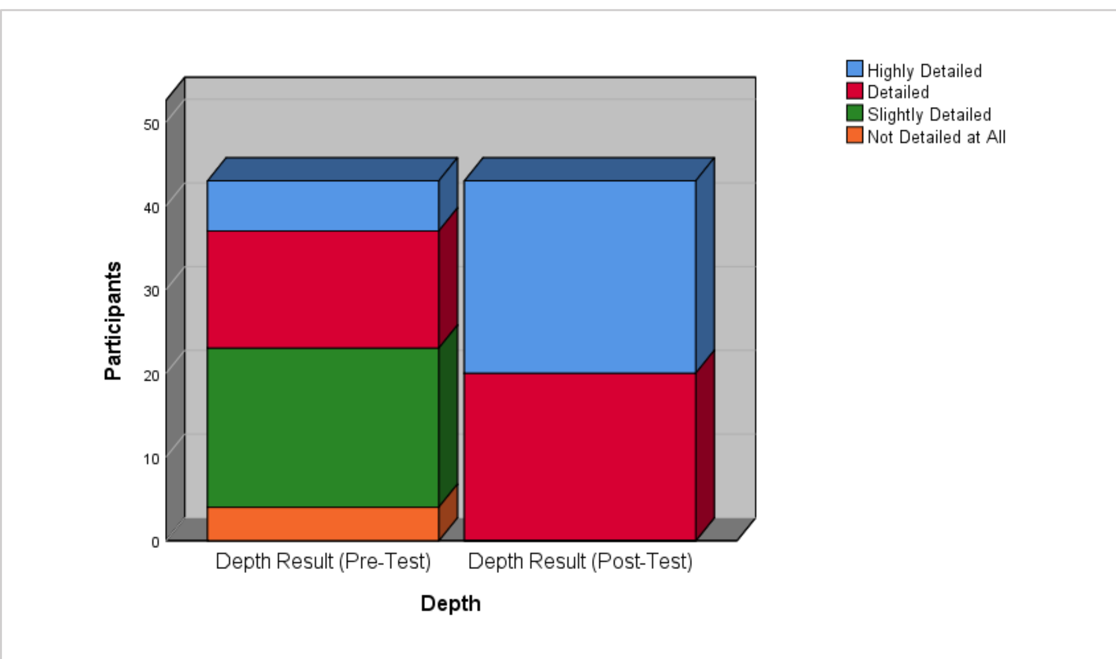

Figure 3. *Results of output depth before and after using role prompting*

The second construct has known a noticeable disparity in responses. Figure 3 presents the result of ChatGPT's generated output depth before and after using role-based technique. Results of pre-test show that a majority of 44.2% found the GPT's response to be Slightly detailed, followed by 32.6% who described the output to be detailed, 14% described it as highly detailed, whereas 9.3% reported the output to be not detailed at all. Results of post-test which aimed at measuring the output depth after using role-based prompting technique show a remarkable difference of 53.5% and 46.5% who described the output to be highly detailed and detailed representing 100% of the participants. No one has found the output in this second part of the experiment to be slightly detailed or not detailed at all.

Wilcoxon Signed-Rank Test confirmed the presence of a strongly significant difference between ChatGPT's depth results of pre-test and post-test with $|z| = 5.72$ and $p < 0.001$.

Following the same procedure explained in the previous section dealing with results of GPT's output clarity before and after the intervention, and in spite of its high significance according to SPSS generated statistics, we decided to extract the exact small p value using python. We got the following result:

$$p = 2.133281006266547e\text{-}08$$

$$p = 0.000000021332$$

Statistical evidence confirms the high significant difference existing between results of the generated output's depth before and after using role-based prompting technique.

After proving the existing difference between depth scores of the pre-test and post-test, it is crucially essential to highlight the direction of those results. To that aim, we ran the Paired Wilcoxon-Signed Rank test on SPSS to get insights into positive and negative ranks, besides ties.

Table 2. *Wilcoxon Signed-Rank Test Table of Ranks (Output Depth)*

		N	**Mean Rank**	**Sum of Ranks**
DepthPosttest - DepthPretest	Negative Ranks	0	,00	,00
	Positive Ranks	37	19,00	703,00
	Ties	6		
	Total	**43**		

As shown in Table 2, 37 students have remarkably shown an increase between pre-test and post-test depth results. 6 have stuck to the same depth rating in pre-test and post-test. Whereas no one reported a decrease in the generated output depth in the post-test.

We conclude that statistical evidence reveals a strong rejection of the null hypothesis for output quality in terms of depth.

4.1.3. Professionalism

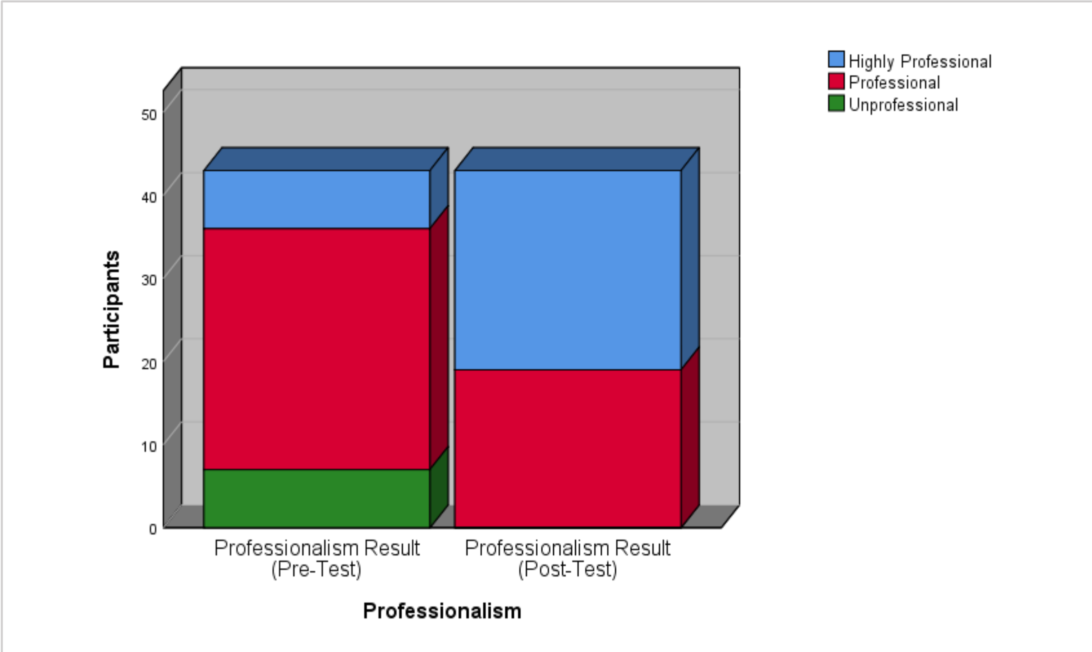

Figure 4. *Results of output professionalism before and after using role prompting*

Figure 4 presents result of the generated output's professionalism before and after using role-based prompting technique. Results of pre-test show that the response generated without using role-based technique was professional according to 67.4% of informants, highly professional according to 16.3%, and unprofessional according to 16.3% of participants. No one has ticked 'Extremely unprofessional' as output generally does not violate basic formality standards. In contrast, results of the post-test show that GPT's generated response to the prompt that used role-based technique was highly professional according to a majority of 55.8% and professional according to 44.2%. The post-test has known the disappearance of claims describing GPT's response as unprofessional and a high increase in numbers of informants who have described the response to be highly professional (7 participants in the pre-test and 24 participants in the post-test).

Results of Wilcoxon Signed Rank Test show a significant difference in the ChatGPT's generated output professionalism rating between pre-test (Mdn=2) and post-test (Mdn=3), $|z| = 4.89$, $p<0.001$ Wilcoxon Signed-Rank Test conducted in google Colab using Python show the exact value of p as follows:

$$p= 9.58400566647125e\text{-}06$$

$$p= 0.000009584$$

Table 3. *Wilcoxon Signed-Rank Test Table of Ranks (Output Professionalism)*

		N	Mean Rank	Sum of Ranks
ProfessionalismPosttest - ProfessionalismPretest	Negative Ranks	0	,00	,00
	Positive Ranks	24	12,50	300,00
	Ties	19		
	Total	**43**		

Table 3 shows the distribution of negative and positive ranks besides ties among participants for the GPT's generated output professionalism rating. Results of GPT's output professionalism have a shown significant difference between pre-test and post-test scores. However, it is mandatory to indicate the direction of the proven difference and potential increase or decrease in the generated output's professionalism. As shown in the table of ranks, the output's level of professionalism has known a significant increase (24 positive ranks and 0 negative ranks). 19 participants have not given any different rating in the post-test compared to their pre-test output professionalism result. No has reported a decrease in GPT's degree of professionalism.

We conclude that statistical evidence suggests a light rejection of the null hypothesis for output quality in terms of professionalism.

4.1.4. Insightfulness

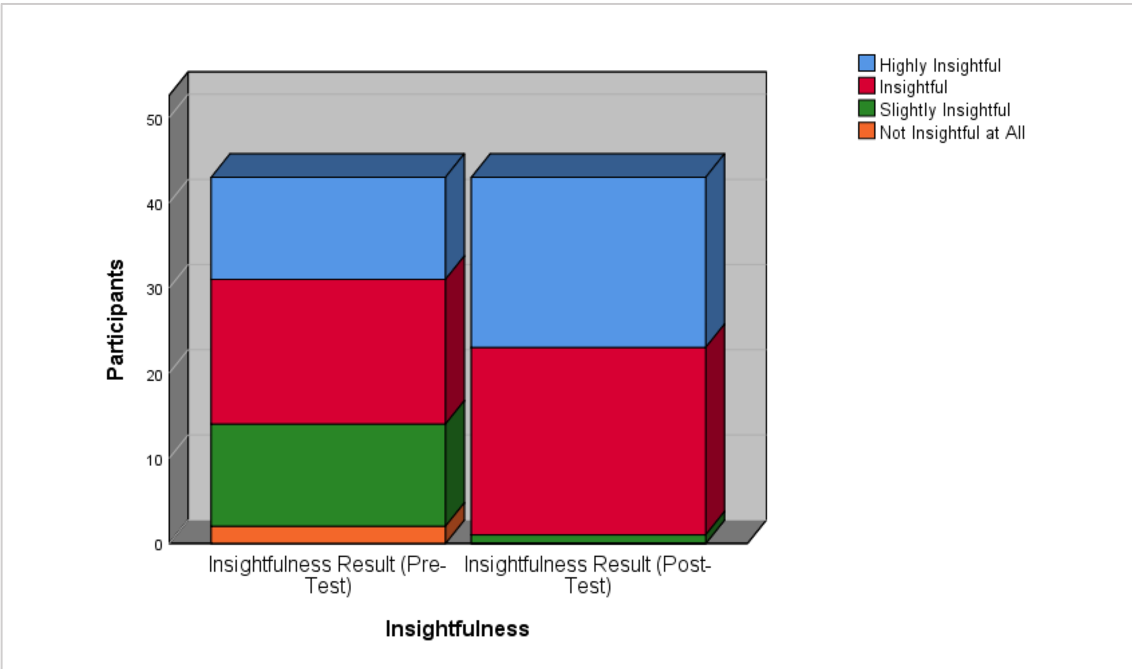

Figure 5. *Results of output insightfulness before and after using role prompting*

Figure 5 shows results of the generated output's insightfulness before and after using role-based technique. Results of the pre-test show that 39.5% participants found the output before using role-based prompt to be insightful, 27.9% found the output slightly insightful, 27.9% described it as highly insightful, whereas 4.7% found that the generated output was not insightful at all. In contrast, results of the post-test which measured output's insightfulness after including role-based prompt show a significant increase as 51.2% found the output insightful, 46.5% found it highly insightful, and only 2.3% reported it to be slightly insightful. Contrary to the pre-test, in the post-test no one strictly denied the insightfulness of the generated output.

Using SPSS, we conducted Wilcoxon Signed-Rank Test and it revealed a high significant difference between output insightfulness results of pre-test and post-test with $|z| = 4.23$ and $p < 0.001$.

Following the same procedure, Wilcoxon Signed-Rank Test was re-conducted on Google Colab using Python, the following result was achieved.

$$p= 1.7783648270427474e\text{-}05$$

$$p= 0.0000177836$$

Table 4. *Wilcoxon Signed-Rank Test Table of Ranks (Output Insightfulness)*

		N	Mean Rank	Sum of Ranks
InsightfulnessPosttest - InsightfulnessPretest	Negative Ranks	0	,00	,00
	Positive Ranks	20	10,50	210,00
	Ties	23		
	Total	**43**		

As shown in Table 4, Wilcoxon Signed-Rank Test conducted using SPSS show a significant increase in ChatGPT's generated output insightfulness between pre-test and post-test (Positive ranks = 20 and Negative Ranks = 0). 23 participants have not reported any change in output insightfulness between pre-test and post-test.

To conclude, results revealed a strong rejection of the null hypothesis for output quality in terms of insightfulness.

4.1.5. Innovativeness

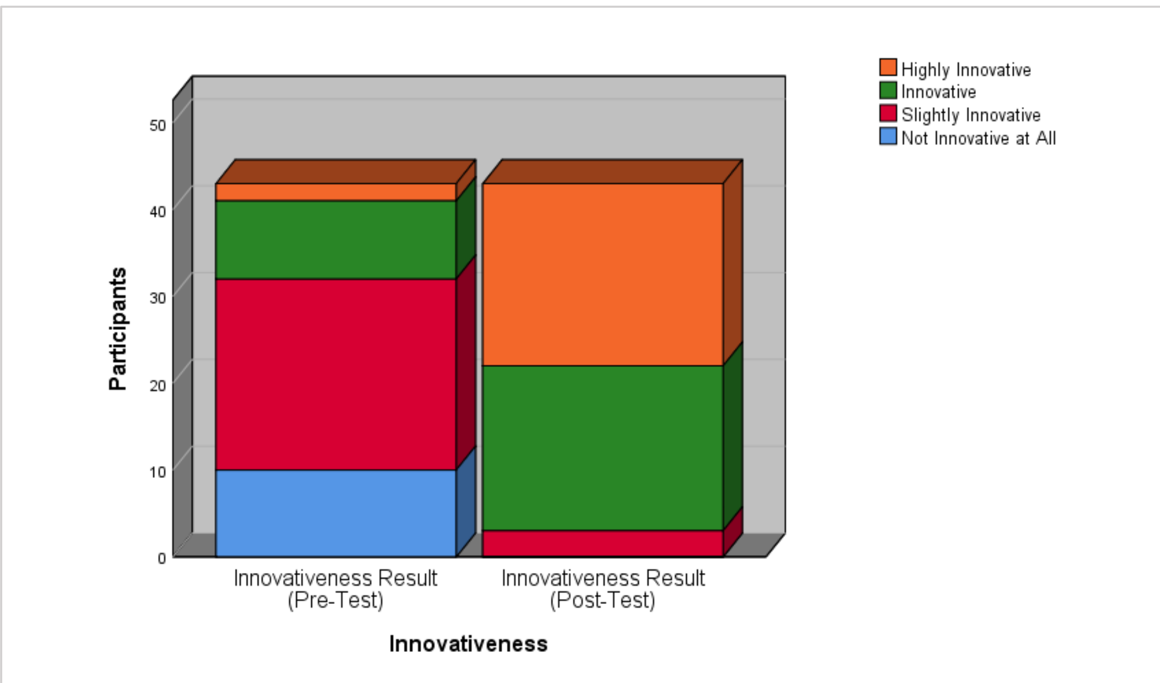

Figure 6. *Results of output innovativeness before and after using role prompting*

Figure 6 shows results of ChatGPT's innovativeness before and after using role-based prompts. Results of pre-test in which role-based technique was not used show that the generated response was slightly innovative according to a majority of 51.2% of participants, 23.3% of participants who described it to be not innovative at all, 20.9% claimed that the output was innovative, whereas only 4.7% found the initial response highly innovative. Results of post-test show significant contrast in innovativeness quality measurement outcome compared to post-test results. According to a majority of 48.8% found the generated output after using role-based prompting highly innovative, followed by 44.2% who described it as innovative, whereas only 7.0% reported that the generated output in the second part of the experiment was slightly innovative. Contrary to initial results, no one denied the presence of innovation in the generated output.

Using SPSS, we conducted Wilcoxon Signed-Rank Test and it revealed a highly significant difference between output innovativeness results of pre-test and post-test with $|z| = 5.78$ and $p < 0.001$ ($p=0.000$ according to SPSS which displays only three decimal places for p).

Results of Wilcoxon Signed-Rank Test using Python have confirmed the strongly significant difference between input innovativeness in the pre-test and post-test with $p < 0,001$ ($p=0.0000000105$).

Table 5. *Wilcoxon Signed-Rank Test Table of Ranks (Output Innovativeness)*

		N	Mean Rank	Sum of Ranks
InnovativenessPosttest - InnovativenessPretest	Negative Ranks	0	,00	,00
	Positive Ranks	41	21,00	861,00
	Ties	2		
	Total	**43**		

Table 5 shows additional statistical details provided by the paired Wilcoxon Test showing the distribution of positive ranks, negative ranks, and ties associated with results of the generated output innovativeness essential to identify the direction of results.

Results of the generated output's innovativeness have known a remarkable increase suggesting the existing impact of role-based prompting on the innovativeness of the generated response of ChatGPT, as shown in the table which displays number of positive ranks (41) representing the number of participants who have witnessed an increase in the degree of innovativeness after using role-based prompting technique, negative ranks (0), and ties (2) representing the number of participants who have kept the same position in both pre-test and post-test without reporting any change in output innovativeness. To conclude, based on statistical evidence we reject the null hypothesis for output quality in terms of innovativeness.

4.1.6. Relevance

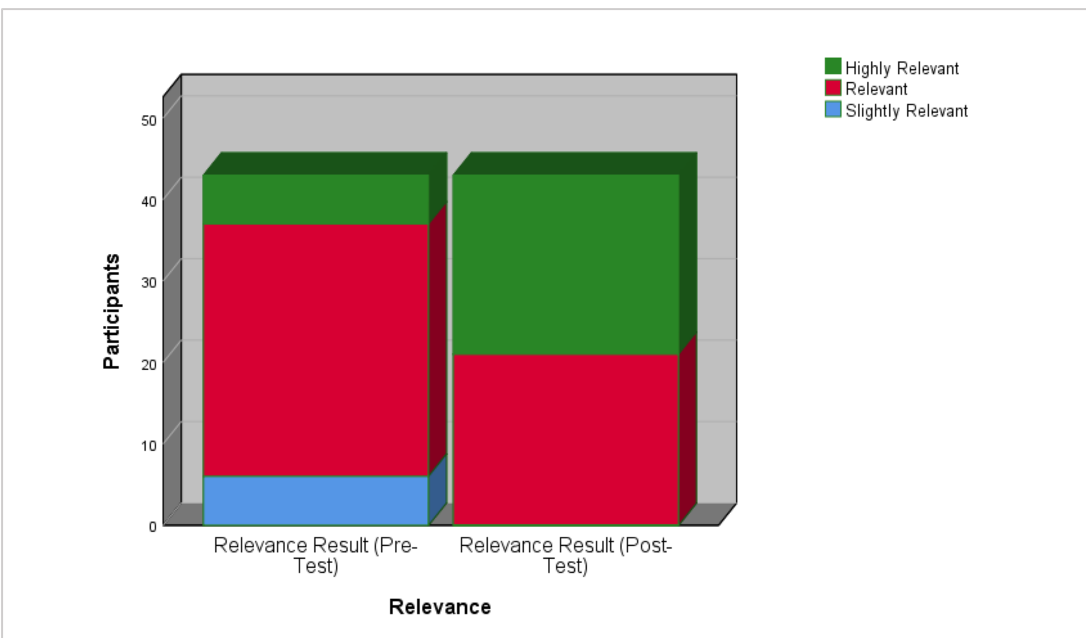

Figure 7. *Results of output relevance before and after using role prompting*

Figure 7 shows results of output quality in terms of relevance before and after using role-based prompts according students' rating. It was found that the initial generated response was relevant according to 72.1% participants, slightly relevant according to 14.0%, and highly relevant for 14,0% of students. No one mentioned that the generated response was not relevant at all as ChatGPT usually provides relevant or slightly relevant responses if prompts are well-formed even without using role-based or other prompting techniques. Output quality in terms of relevance has known a remarkable difference after using role-based prompts as results show that ChatGPT's generated output was highly relevant for 51.2% of participants and relevant for 48.8% of students. Additionally, there was a noticeable shift in results characterized by the disappearance of claims that described the output as slightly relevant in the pre-test.

We initially conducted the paired Wilcoxon Signed-Rank Test on SPSS and it revealed that changes between pre-test and post-test results are statistically significant with $|z| = 4.69$ and $p < 0.001$ ($p=0.000$ according to SPSS which reports p with only 3 decimal places). To overcome this and as we did for the rest of variable paired results between pre-test and post-test, we extracted the Paired Wilcoxon Test exact small value of p using Python on Google Colab coding environment, results have confirmed the previous interpretation as:

$$p= 3.215955563200678e\text{-}05$$

$$p= 0.00003215$$

Hence, we safely conclude that $p < 0.05$ based on the strong statistical evidence we have reached. Thus, we reject the null hypothesis for the generated output relevance.

For statistical clarifications on ranks which demonstrate the direction of results, Wilcoxon Signed-Rank Test's following table shows the distribution of positive ranks, negative ranks, besides ties of output relevance results.

Table 6. *Wilcoxon Signed-Rank Test Table of Ranks (Output Relevance)*

		N	**Mean Rank**	**Sum of Ranks**
RelevancePosttest - RelevancePretest	Negative Ranks	0	,00	,00
	Positive Ranks	22	11,50	253,00
	Ties	21		
	Total	**43**		

As shown in Table 6, 22 students have reported an existing difference between output relevance before and after using prompting technique, no one has reported a decrease in relevance during post-test compared to pre-test. However, 21 students have not noticed an increase or decrease in output relevance after using prompting technique, keeping the same position.

To conclude, statistical evidence proved an existing increase in the generated response's relevance after using role prompting technique, revealing a strong rejection of the null hypothesis for output quality in terms of relevance.

4.1.7. Generosity

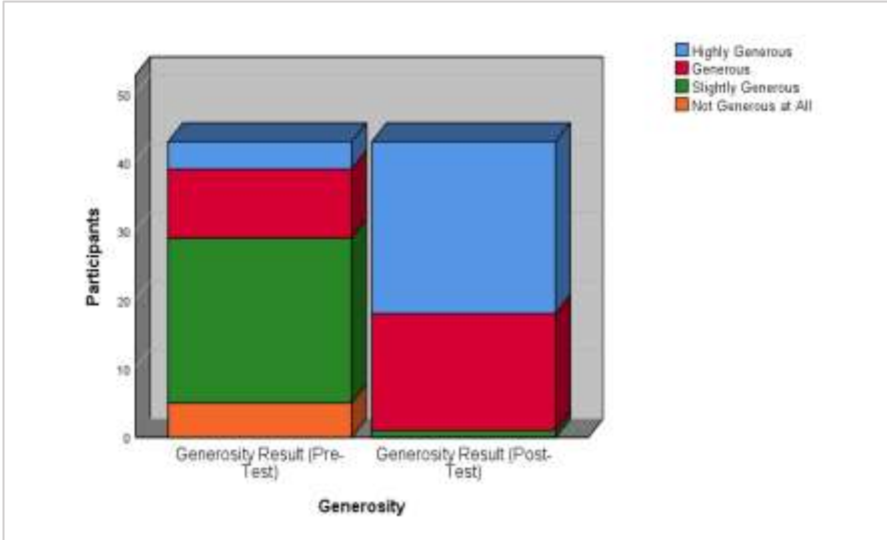

Figure 8. *Results of output generosity before and after using role prompting*

As shown in Figure 8, we asked students to rate the output quality in terms of generosity before using role-based prompting technique. It was found that it was slightly generous according to 55.8% participants, generous according to 23.3%, not generous at all according to 11.6%, and highly generous for 9.3% of participants. Results of post-test show a significant increase in quality more specifically in terms of generosity after using the role-based prompt. 58.1% of participants mentioned it to be highly generous (Compared to 9.3% in the pretest), 39.5% mentioned that it was generous (compared to 23.3%), and slightly generous for 2,3% (compared to 55.8% in the pre-test).

Following the same path, we conducted the paired Wilcoxon Signed-Rank Test on SPSS which revealed a strongly significant difference in output generosity between pre-test and post-test results with $|z| = 5.68$ and $p < 0{,}001$ (p=0.000 according to SPSS which displays p with only three decimal places)

Subsequently, we ran the paired Wilcoxon Signed-Rank Test using Python code on Google Colab to extract the exact p value for more statistical transparency. Results have confirmed the strong significant difference between both results with:

$$p= 1.8812637634480933e\text{-}08$$

$$p= 0.00000001881$$

Now that we have confirmed the existing difference between pre-test and post-test results of output generosity, it is very important to provide evidence on the potential increase in output generosity in post-test results compared to those of pre-test. We ran the paired Wilcoxon signed-rank test again and the following table provide statistical insights into positive ranks, negative ranks, and ties.

Table 7. *Wilcoxon Signed-Rank Test Table of Ranks (Output Generosity)*

		N	Mean Rank	Sum of Ranks
GenerosityPosttest - GenerosityPretest	Negative Ranks	0	,00	,00
	Positive Ranks	39	20,00	780,00
	Ties	4		
	Total	**43**		

As shown in table 7 which gives us a further insight into the direction of the difference between both results. 39 students have reported an improvement in the output generosity after using role-based prompting technique (Positive Ranks), 4 have kept the same position as they did not report the output to be more generous or rich in information after using role-based technique compared to the initial one (Ties). Whereas no one indicated a decrease in the output generosity after using the prompting technique (Negative Ranks).

To conclude, statistical evidence suggests a strong rejection of the null hypothesis for output generosity.

4.2. Satisfaction

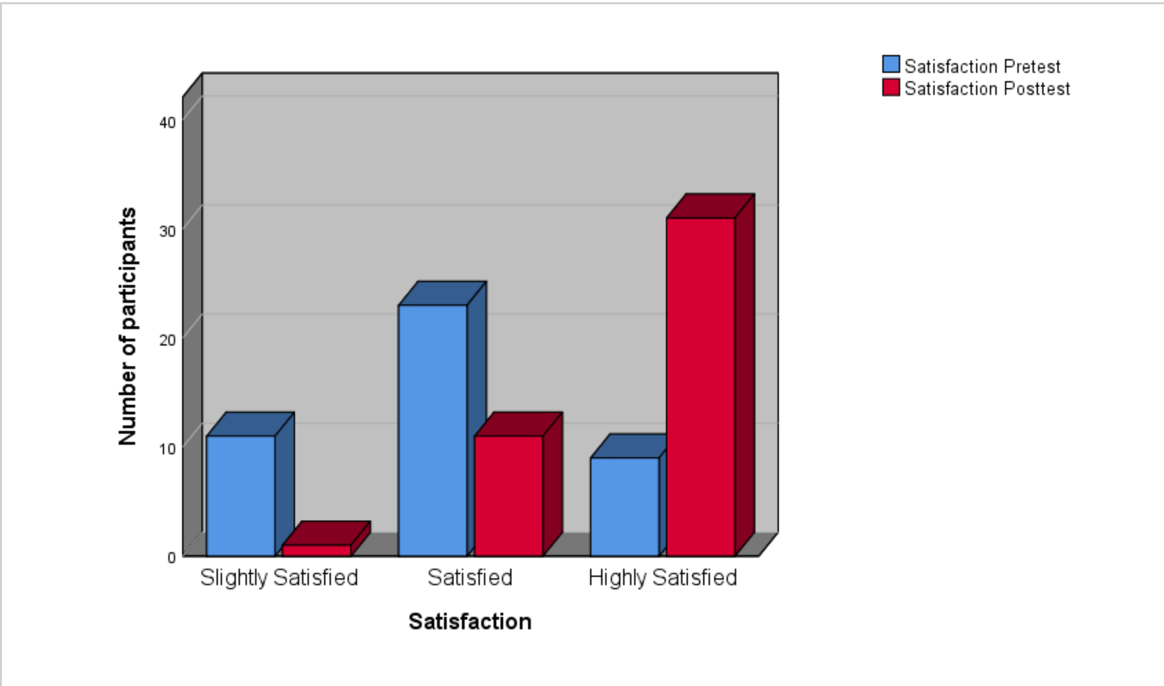

Figure 9. *Results of students' satisfaction with the output quality before and after using role prompting*

As shown in Figure 9 which presents results of students' overall satisfaction with the ChatGPT's generated output before and after using role-based prompting. 72.1% have expressed that they were highly satisfied in the post-test compared to the pre-test result 20.9%. 25.6% claimed that they were satisfied in the post-test compared to 53.3% in the pre-test. 2.3% claimed they were slightly satisfied with the output quality in the post-test compared to 25.6% in the pre-test. It is to note that the question aiming at measuring satisfaction was measured at 4 Point-Likert scale. No one has expressed their strict overall dissatisfaction with the generated output in pre-test or post-test.

To test the significance of the noticed difference, we conducted the paired Wilcoxon Signed-Rank Test on SPSS which revealed a strong significant difference between students' satisfaction rates in the pre-test and post-test $|z| = 5.1$. As SPSS limits p value to three decimal places and due to its small value, we got $p < 0.001$ ($p = 0.000$).

To strengthen our results, we decided to overcome this limitation by running the paired Wilcoxon Signed-Rank Test using Python code on Google Colab to obtain the exact p value. Statistical results have confirmed the significant difference in satisfaction rates between pre-test and post-test.

$$p = 1.1388565071396583e\text{-}06$$

$$p = 0.0000011388$$

Now that we have had strong statistical evidence proving the existence of a significant difference between satisfaction results before and after using role-based AI-prompting. It is very important to shed light on the direction of those results to identify whether satisfaction has increased or decreased after researchers' intervention. To this aim, paired Wilcoxon Signed-rank Test provides insights in positive and negative ranks besides ties as shown in the table below.

Table 8. *Wilcoxon Signed-Rank Test Table of Ranks (Satisfaction)*

		N	Mean Rank	Sum of Ranks
SatisfactionPosttest - SatisfactionPretest	Negative Ranks	0	,00	,00
	Positive Ranks	29	15,00	435,00
	Ties	14		
	Total	**43**		

Table 8 shows that there was a noticeable increase in students' satisfaction with the overall quality of the generated output by ChatGPT after using role-based prompting compared to the initial output with 29 students who have shown an increase in satisfaction besides 14 students who have not shown any difference in their satisfaction between pre-test and post-test. Whereas no one has shown a decrease in their satisfaction in the post-test results.

We conclude that there is a significant increase in students' overall satisfaction in the post-test results. Statistical evidence suggests a strong rejection of the null hypothesis for students' satisfaction.

5. Discussion

To answer two main questions raised at the beginning of this paper, results' section was split into two parts: Output Quality and Satisfaction. Each part presented results adequate to come up with an answer to the appropriate question. Output quality before and after using role-based technique was measured based on students' evaluations using 7 main criteria.

For output clarity, it was found that using role-based prompts enhances the generated content's clarity as there was a noticeable difference between output quality in terms of clarity between the initial output in which role prompts were not used and the second output which was reportedly clearer after using role-based technique.

For output depth, it was found that responses generated after using role-based technique were more detailed compared to the previous ones, suggesting that role-based prompts help students extract more detailed answers from ChatGPT depending on their need compared to ordinary prompts which are not based on any technique mainly the role-based one.

For output professionalism, it was found that there is a slight increase in the degree of professionalism when comparing the two outputs generated with and without using role-based prompts. This is explainable as ChatGPT's professional responses are not strictly tied to using specific technique as the chatbot model tends to act professional in the majority of cases especially when used for learning, following basic universal professionalism standards. The main difference here lay in results moving from professional to highly professional between pre-test and post-test results, suggesting an improvement in the degree of professionalism without denying the initial degree of output professionalism noticed even before using the role-based technique.

For output insightfulness, it was found that the generated responses were more insightful after using role-based prompts compared to the initial ones. Students' centeredness here was extremely essential as we got to rate insightfulness of the output based on students' understanding to assess the importance of the prompting technique in AI-supported learning.

For output innovativeness, results have demonstrated that the use of role-based prompts can leverage the output's innovativeness which positively impacts the overall quality of ChatGPT's responses contributing to more effective assistance.

For output relevance, students have reported the output after using role-based prompts to be more relevant compared to the initial output generated before using the technique in question. This does not deny the relevance of outputs using good inputs which do not assign the chatbot a specific role. However, role-based prompting technique can, according to our findings, make a chatbot generate comparatively more relevant answers.

The last element we measured for quality assessment was the output's generosity. Students described the response they got after using role-based prompts as richer in information and more generous compared to the initial one highlighting the importance of role-based prompts in improving the overall quality of ChatGPT's outputs.

To answer the second question, the second part of this paper's results dealt with students' satisfaction scores between pre-test and post-test. Strong statistical evidence revealed the significant difference between students' overall satisfaction with the generated output before and after using role-based prompting (results of pre-test and post-test). More specifically, students' satisfaction has significantly increased as post-test results have shown, suggesting the importance role-based prompts can have in this respect.

In simple words, Rosas et al. (2024) highlight, "Chatbots are computer programs capable of understanding and responding to human language." (p. 115). For although chatbots' use does not require technical expertise in computer science as these technologies process and use natural language, rendering the human-chatbot interactions smoother, AI prompting as a skill remains a fundamental component of chatbot's use.

6. Conclusion

The present quasi-experimental study aimed at investigating role-based prompting technique and its role in leveraging AL models' output quality in AI-assisted learning following a student-centered approach. Results and statistical evidence revealed a strongly significant difference between the selected model's generated output quality with and without using role-based prompts. Additionally, it was found that role-based prompting raises AI models potential positively contributing to the process of learning using generative AI. The present paper also sheds light on the game-changing role this technique plays in not only taking chatbots' responses to the next level in terms of clarity, depth, professionalism, insightfulness, innovativeness, relevance, and generosity, but also in raising students' overall satisfaction with the generated output quality.

Limitations of this study include the relatively small sample size as larger samples might allow us to safely reduce the impact of students' potential lack of accurate evaluation abilities. Future studies should evaluate students' awareness of AI-prompting skills as many learners who do not specialize in computer science tend to have little to no knowledge about the importance of prompting in AI-assisted learning. Further studies should also investigate the possibility of integrating trainings to develop students' skills in using generative AI for a positively impactful use.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Bozkurt, A., & Sharma, R. C. (2023). Generative AI and prompt engineering: The art of whispering to let the genie out of the algorithmic world. Asian Journal of Distance Education, 18(2), i-vii. **https://doi.org/10.5281/zenodo.8174941**

[2] Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative ai. *Business & Information Systems Engineering*, *66*(1), 111-126. **https://doi.org/10.1007/s12599-023-00834-7**

[3] Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Education Sciences*, *13*(9), 856. **https://doi.org/10.3390/educsci13090856**

[4] Mittal, U., Sai, S., & Chamola, V. (2024). A comprehensive review on generative ai for education. *IEEE Access*. **https://doi.org/10.1109/ACCESS.2024.3468368**

[5] Oppenlaender, J., Linder, R., & Silvennoinen, J. (2024). Prompting AI art: An investigation into the creative skill of prompt engineering. *International journal of human–computer interaction*, 1-23. **https://doi.org/10.1080/10447318.2024.2431761**

[6] Park, J., & Choo, S. (2024). Generative AI prompt engineering for educators: Practical strategies. *Journal of Special Education Technology*, 01626434241298954. **https://doi.org/10.1177/01626434241298954**

[7] Rosas, R. A. P., & Rodríguez, E. J. A. (2024). Chatbots: The Future of Education? *International Journal of Engineering Pedagogy*, *14*(8). **https://doi.org/10.3991/ijep.v14i8.49715**

[8] Theophilou, E., Koyutürk, C., Yavari, M., Bursic, S., Donabauer, G., Telari, A., ... & Ognibene, D. (2023, November). Learning to prompt in the classroom to understand AI limits: a pilot study. In *International Conference of the Italian Association for Artificial Intelligence* (pp. 481-496). Cham: Springer Nature Switzerland. **https://doi.org/10.1007/978-3-031-47546-7_33**