

Chapter

Influence of Prompts Structure on the Perception and Enhancement of Learning through LLMs in Online Educational Contexts

Silvia Rodriguez-Donaire

Abstract

This research examines how the structure of prompts impacts the perceived depth and accuracy of responses generated by generative Large Language Models (LLMs) in educational settings. It specifically investigates how prompt design influences students' learning experiences. The study involved an experiment with 183 students enrolled in a mandatory Business Administration course at the Universitat Oberta de Catalunya (UOC). Data from the experiment were analyzed using both qualitative and quantitative methods. The results show that well-structured prompts significantly improve students' perception of the depth and accuracy of GenAI-generated responses, leading to a more effective learning process. This underscores the crucial role of prompt design in maximizing the educational effectiveness of GenAI. The findings suggest that thoughtful prompt design can enhance educational outcomes, although the study's limited sample size and context-specific nature may restrict the generalizability of the results. This research contributes to the field by highlighting the importance of prompt structure in harnessing GenAI tools for educational improvement.

Keywords: generative AI, GenAI, engineering prompts, large language model, LLMs, formative assessment, online higher education, prompt structure, educational environment, students' perception, learning process

1. Introduction

Artificial intelligence (AI) tools such as OpenAI's ChatGPT, Google's Gemini, and Microsoft's CoPilot are changing how we perform everyday tasks, from writing emails to creating multimedia content. However, the use of AI in education presents new challenges and opportunities, necessitating a re-evaluation of the current educational model. This study aims two-fold: (1) pinpoint the elements of prompt structure that affect the quality of GenAI-generated responses, and (2) examine

how prompt structure relates to the perceived effectiveness of GenAI responses in improving the learning experience in educational environments. It is worth mentioning that the structure of prompts is crucial in determining the quality of AI-generated responses as it directly impacts the model's ability to understand and respond appropriately. For instance, [1] study demonstrated that modifying the queries can enhance the performance of language models by 11.46%, indicating that prompt structure significantly affects response accuracy. In addition, [2] emphasized the effectiveness of ChatGPT in structured tasks, but also pointed out its limitations in more nuanced applications, underscoring the importance of prompt structure in optimizing AI performance.

On the one hand, specific elements of the prompt structure influence the quality of generated responses. Bozkurt [3] underlines the significance of prompt engineering as a form of digital literacy, vital for utilizing AI in education. Liu [4] research has identified crucial components in the generated responses, such as definitions and examples, which are essential for the accuracy and reliability of the information.

On the other hand, prompt engineering can greatly enhance the quality of educational applications [5]. For example, techniques like Chain-of-Thought and Ask-me-Anything have improved the quality of answers in mathematical problem-solving [6]. However, despite these advancements, challenges still exist in effectively implementing and validating AI tools in educational settings. This includes the need for AI literacy and the development of prompt engineering skills [7, 8].

Significant progress has been made in understanding how the structure of a prompt influences the accuracy and quality of AI-generated responses in the educational context. However, there are still research gaps that need to be addressed. While current studies have covered general aspects of prompt engineering, further investigation is necessary to explore the specific elements of prompt structure that affect different types of AI-generated responses in various educational contexts [1, 2]. Additionally, more research is needed to understand how these prompt engineering techniques can be effectively integrated into educational settings to improve the perception and effectiveness of AI in the learning process [5, 6]. Additionally, it is essential to explore the ethical implications and challenges associated with AI implementation in education, particularly in terms of ensuring accuracy and addressing bias [3, 7].

Alternative approaches exist, such as integrating frameworks like the CLEAR Framework (Concise, Logical, Explicit, Adaptive, and Reflective) to optimize interactions with AI language models [9]. Other approaches include implementing active methodologies and providing digital literacy training for educators and students, which could offer additional solutions to improve the effectiveness of AI in education [10, 11].

This study investigates how prompt design can enhance the accuracy and depth of GenAI-generated responses in online educational settings and its impact on students' learning processes. As a result, the research aims to answer the following questions:

1. How does the structure of prompts affect the perceived accuracy of GenAI responses in an educational setting?
2. What specific elements of prompt structure influence the quality of GenAI responses?

3. What is the relationship between the prompt's structure and the AI's perceived effectiveness in enhancing the learning process?

This research will be conducted through an empirical study and data analysis on the prompts created by students and their perceptions of the responses received from GenAI. To achieve this, a methodological approach will be used, including coding the prompts and analyzing student perception through statistical analysis of the resulting database.

The chapter is structured as follows. First, a literature review on using GenAI in higher education is conducted. Next, the methodology used to carry out the empirical study is explained. Then, the activity that takes place within the online course is described. Finally, the main results found in the experiment are presented and discussed, and the study concludes by identifying the limitations and future research areas.

2. Literature review

The history of artificial intelligence can be traced back to 1854 when George Boole developed it. However, the term “artificial intelligence” was first coined by John McCarthy in 1956 during the Dartmouth Conference, which laid the foundation for its development. Machine learning algorithms were advanced during the 1980s and 1990s, allowing AI to learn and improve autonomously. In 1997, IBM's DeepBlue computer defeated world chess champion Garri Kaspàrov, and iRobot launched Roomba, the first autonomous robot vacuum cleaner, which became a commercial success. The 2010s saw the development of deep neural networks, which enabled AI to perform complex tasks such as image recognition and natural language processing. As a result, virtual assistants such as Apple's Siri, Microsoft's Cortana, and Amazon's Alexa were launched, and generative AI capable of creating content such as text, images, and music emerged. This opened up new creative possibilities, leading to the launch of ChatGPT and other applications that consolidated content creation globally by 2022.

A significant amount of research [12–18] highlights the potential of GenAI to enhance learning and teaching in various knowledge disciplines. However, despite the many perceived benefits of AI, its use in higher education requires an exploration of potential ethical challenges and reservations related to academic integrity. In recent years, the higher education sector has faced different challenges, primarily due to the COVID-19 pandemic. The most significant challenge for most universities has been adapting in-person teaching to remote teaching. This situation has led to a reconsideration of how academic programs are delivered and the introduction of more flexible approaches to learning and teaching within universities.

Integrating generative AI¹ in higher education is a challenge that questions universities' traditional role in knowledge production and dissemination. According to [20], universities worldwide have varied responses to this technological advancement. Some institutions prohibit using generative AI tools due to concerns about academic

¹ Generative AI—“computational systems trained on large data sets to generate human-like responses or outputs” [19].

integrity and the potential hindrance of independent thinking and creativity among students and faculty. On the other hand, these tools can support the generation of ideas and enrich discussions in the context of teaching and learning.

According to [21], AI has several applications in various sectors, and some of these can also be replicated in the education sector. For instance, chatbots can provide 24/7 assistance to students, optimization algorithms can offer personalized learning that adapts to each student's needs, and automating administrative tasks can allow teachers to focus more on teaching.

Although AI has many potential advantages, such as personalization of content, accessibility of information, and promoting more interactive education with immersive and compelling environments, it also poses risks. These include data protection and bias of AI algorithms [22], dehumanization of education, an increase in the digital divide between students with different access to advanced technologies, and excessive reliance on technologies that limit the development of cognitive and problem-solving skills [23]. AI is another technological revolution, like the appearance of the Internet. That is why we must live with it, learn from it, and transform our educational models by training students to use AI ethically and responsibly.

Interacting with LLMs, such as ChatGPT, Gemini, Copilot, etc., might seem easy initially, but it requires a certain level of digital skills and knowledge to use them effectively. Teaching students how to create clear and concise instructions for LLMs is crucial, as it is usually a trial-and-error process [24]. This is because the quality of the instructions directly affects the results' quality. According to [25], a well-crafted prompt can lead to a precise, accurate, and relevant response from the LLMs, thereby maximizing its performance. Conversely, a poorly structured question can result in an ambiguous, incorrect, or irrelevant answer.

To make the most of the capabilities of LLMs, it is essential to master the art and science of formulating effective prompts, known as "prompt engineering". This requires a combination of domain-specific knowledge, model understanding, and skills that can only be honed through experience and learning. By equipping students with these skills, they can use AI safely, ethically, and effectively, receive more accurate answers, understand the limitations of LLMs, and develop new knowledge and skills to adapt to an ever-changing environment.

3. Methodology

This research is part of a qualitative study that explores the relationship between the effective use of prompts in generative AI and the quality and accuracy of the response obtained in a specific activity. The study also aims to determine if this approach positively impacts the participants' learning process. More specifically, the research focuses on a debate activity that is part of a mandatory subject called "Information Systems" (IS) taught in Catalan and Spanish. The subject is offered as part of the curriculum for the Business Administration and Management (ADE) degree at the Universitat Oberta de Catalunya (UOC). This entirely online university follows an asynchronous educational model.

The study was conducted on 304 students enrolled in four Spanish and two Catalan classrooms. The Spanish classrooms had a maximum capacity of 55 students, while the Catalan classrooms had 43 students. Out of these 304 students, only 194

participated in the debate activity, where they were required to give feedback to one of their classmates with the help of an LLM. Further analysis revealed that only 182 students had included the prompt used in their debate response and had reflected on the response given to them by the LLMs. These twelve values were discarded in the study. To ensure the confidentiality of the participants, all student data has been anonymized.

In each classroom, the students' responses from a discussion exercise were manually coded into binary variables with the assistance of the generative AI ChatGPT 4.0. One instruction was generated to evaluate the effectiveness of prompts created by students within a debate activity, and a second instruction was used to evaluate the students' perception of the responses generated by the LLMs. Both instructions can be found in an open repository [26].

Before proceeding to the quantitative analysis, the coding carried out by Generative AI is reviewed. It is worth noting that in some cases, the coding is not completely correct, and the study authors make manual modifications. Once the data was reviewed, a combination of quantitative and qualitative analysis was applied to confirm the experimental model.

Table 1 shows the descriptive analysis of the final sample of the present article: The distribution of the categorical variables is shown in **Figure 1a–c**.

Variable	Mean	Median	SD	SE	Kurtosis	Skewness
ROLE	0.4250000	0	0.4964157	0.0453163	1.092072	0.3034330
OBJETIVES	1.0000000	1	0.0000000	0.0000000	NaN	NaN
CONTEXT	1.0000000	1	0.0000000	0.0000000	NaN	NaN
SPECIFIC	1.0000000	1	0.0000000	0.0000000	NaN	NaN
VERACITY	1.0000000	1	0.0000000	0.0000000	NaN	NaN
DEPTH	0.3000000	0	0.4601790	0.0420084	1.761905	0.8728716
LEARNING	0.7416667	1	0.4395535	0.0401256	2.219282	-1.1042112
AI_Mark	7.9333333	9	3.2866763	0.3000311	1.895030	-0.4784266
EFFECTIVENESS	3.4250000	3	0.4964157	0.0453163	1.092072	0.3034330

Table 1.
 Descriptive analysis of the study sample.

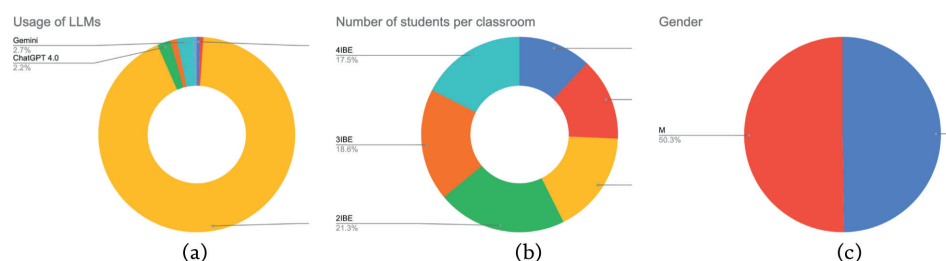


Figure 1.
 (a) Typology of LLMs used, (b) number of students per classroom and (c) sample by gender.

4. Experiment design

This project aims to explore using an LLM such as ChatGPT, Gemini, or Copilot as a support tool for students in the virtual classroom. The project proposes that students provide feedback on an initial intervention from one of their peers through the debate area. The UOC's virtual platform, Canvas, assigns this intervention randomly.

As **Figure 2** shows, the activity requires students to write a prompt in a generative AI that identifies the strengths and areas for improvement in the assigned answer, along with recommendations to improve the weaknesses. As part of the exercise, students are expected to verify the accuracy and depth of the AI-generated information and



Your intermediate intervention in response to the initial intervention of the assigned partner should include the following:

1. The Generative AI used in the exercise.
2. The prompt used (questions you ask within generative AI). Within your prompt, you should include the initial intervention that has been evaluated. To keep your response brief, you can replace this information with the phrase "[Initial intervention of the student's name and last name]"
3. The answer that generative AI provides you.
4. Reason truthfulness and depth of the response proposed by the AI.
5. Make a final evaluation of the AI's response according to the following criteria:
 - **Reliability:** the information provided was accurate and free of inconsistencies or errors.
 - **Quality:** the information provided was of high quality.
 - **Completeness:** the information provided was complete.
 - **Originality:** the information provided was original and not copied from other sources.
 - **Creativity:** the information provided showed creativity.
 - **Multiple perspectives:** the information provided contained different points of view that added value.
 - **Coherence:** the information provided was consistent and logical.
 - **Conceptual relevance:** the information provided was related to the concepts discussed.
 - **Suitable examples:** the information provided included appropriate examples.
 - **Improvement:** the AI helped you improve your reflection on critical points.

Figure 2.
Statement of the debate activity—Feedback on an Intervention.

	EXCELLENT (10)	GREAT (9)	GOOD (7)	PASS (5)	FAIL (3)	DEFICIENT (0)
Argumentative intervention Reasoning and reflection of the AI response (20 Points out of 50 points)	Reason and thoroughly reflect on the response provided by the AI in relation to the interventions of classmates, with relevant, well-developed, justified contributions based on concepts studied in modules and/or valuable examples.	Reason and reflect adequately on the responses of the AI provided in relation to the interventions of your peers, making relevant, developed, justified, and concept-based contributions impact on the concepts of the modules.	Reason and minimally reflect on the responses provided by the AI in relation to the interventions of your peers, with sufficient but undeveloped contributions and without examples or with little impact on the concepts of the modules.	Insufficiently reasons and reflects on the responses provided by the AI in relation to the interventions of their peers, with insufficiently argued and/or developed contributions.	Carry out superficial interventions, without exposing the response proposed by the AI, or reasoning or reflecting on the AI's response in relation to the interventions of your peers, and without providing any argumentation.	Does not make argumentative interventions.
Correct use of AI within the argumentative intervention (5 out of 50 points)	Correctly identify the generative AI used, the prompt used in the generative AI, as well as the response proposed by it.	Identify the generative AI used, part of the 'prompt' used in the generative AI, as well as part of the response proposed by it.	Identify the generative AI used, part of the 'prompt' used in the generative AI, as well as part of the response proposed by it.	Identify the generative AI used, part of the 'prompt' used in the generative AI, as well as part of the response proposed by it.	The generative AI used, the prompt, and the response proposed by the AI are not identified.	

Figure 3.
Criteria that teachers evaluate on the part of the activity that involves the use of AI.

complement it with their own analysis and perspective. Finally, at the end of the activity, students are asked to evaluate the answer provided by the AI based on various qualitative variables such as reliability, quality, completeness, originality, creativity, and coherence.

After the debate activity was closed, the students' responses were evaluated using a rubric (see **Figure 3**).

Figure 3 presents the two criteria used to evaluate the activity. The first criterion assesses the proper use of AI in the exercise and carries a weight of 5 points out of 50 points. The second criterion evaluates the reasoning and reflection shown in the response provided by AI. It aims to assess the student's development of critical and analytical thinking skills concerning the application of AI and carries a weight of 20 points out of the total 50 points.

The students' prompt design is also evaluated as part of the experiment. This evaluation aims to determine if the prompts are efficient and whether an efficient design impacts the perceived depth and veracity of the answer provided by the AI. It also seeks to understand whether an efficient prompt design has any relationship with students' perception of the usefulness of AI in the learning process.

Therefore, it is crucial that the prompt follows a clear and concrete structure and includes the parts shown in **Figure 4**.

- Indicate the role or character the LLM must adopt in its response (e.g. acts as...);
- Specify the objective/purpose of the expected response;
- Add relevant information about the context and give examples to specify the answers;
- Be explicit in the instructions: the tone and style of the response, the target to whom it is directed, and expectations, among others.

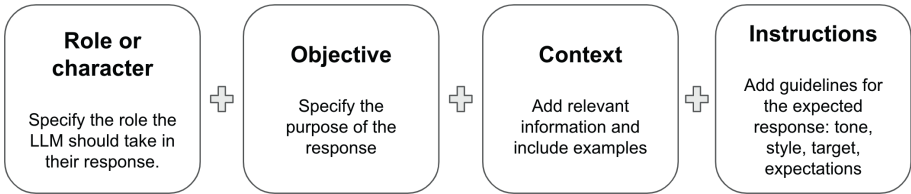


Figure 4.
Parts that an efficient prompt must contain.

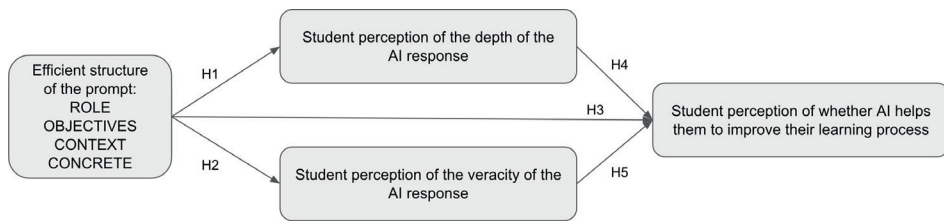


Figure 5.
Experiment model.

A proposed model (**Figure 5**) examines the correlation between the structure of a prompt, the student’s perception of depth, the truthfulness of their answers to the LLMs (Learning Module Materials), and their learning progress.

The model’s hypotheses (**Figure 5**) aim to help us confirm three things: (1) the relationship between the structure of prompt and student’s perception of the answers obtained by the LLMs in terms of depth (H1) and veracity (H2); (2) the relationship between the structure of prompts (H3), students’ perception of depth (H4) and veracity (H5) of answers obtained by the LLMs, and the improvement of the student’s learning process. The hypotheses are detailed below:

- H1—An effective prompt structure contributes positively to the depth perception of the response obtained by the LLMs used.
- H2—An effective prompt structure contributes positively to the truthfulness perception of the response obtained by the LLMs used.
- H3—An effective prompt structure contributes positively to learning enhancement.
- H4—A student’s perception of depth contributes positively to learning improvement.
- H5—A student’s perception of truthfulness contributes positively to learning improvement.

5. Results and discussion

The current analysis evaluates how various prompting styles impact the perception of depth, truthfulness, and learning improvement in LLM responses and how these factors differ among different groups of students. The statistical findings of the comparison between GLM and Random Forest models (**Table 2**) and ANOVA (**Table 3**) and the discussion of the results reinforced by qualitative data obtained in the classrooms are presented below.

A comparison of GLM and Random Forest models (**Table 2**) reveals similar accuracy at 77.78% and AUC of 0.82, indicating overall solid performance. However, differences exist in sensitivity and specificity:

- GLM: Higher specificity (86.67%), lower sensitivity (33.33%).

Modelo	Accuracy	Sensitivity	Specificity	AUC
GLM	0.7777778	0.3333333	0.8666667	0.8222222
Random Forest	0.7777778	1.0000000	0.7333333	0.8222222

Table 2.
Results of the GLM and Random Forest Models.

Variable	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Gender	1	0.0992029	0.0992029	0.8340435	0.3643806
Classroom	5	7.2374421	1.4474884	12.1696883	0.0000000
AI_Type	5	0.6257711	0.1251542	1.0522281	0.3947219
AI_Mark	9	1.5396431	0.1710715	1.4382749	0.1898325
ROLE	1	0.0744499	0.0744499	0.6259338	0.4316426
DEPTH	1	0.5695996	0.5695996	4.7888811	0.0321349
Gender:Classroom	5	0.7050956	0.1410191	1.1856114	0.3256066
Classroom:AI_Type	1	0.1380074	0.1380074	1.1602907	0.2852685
Residuals	67	7.9691215	0.1189421	NA	NA

Table 3.
ANOVA results.

- *Random Forest*: Higher sensitivity (100%), lower specificity (73.33%).

Both models have strengths depending on the problem context. GLM is better at avoiding false positives, while Random Forest is more effective at identifying all positive cases.

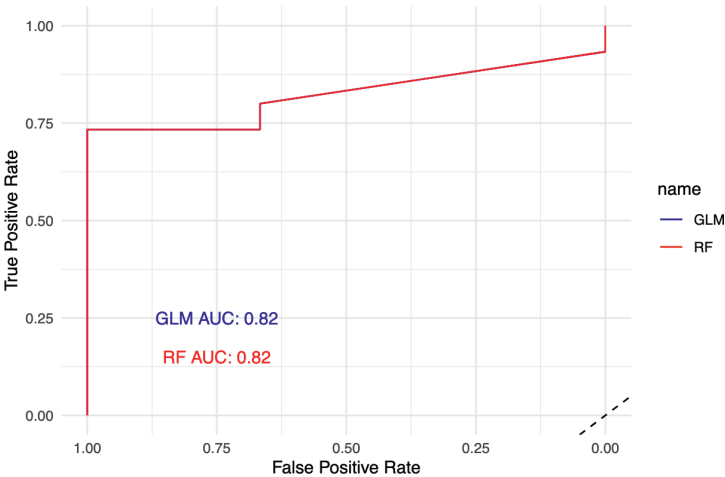


Figure 6.
ROC curves for GLM and Random Forest Models.

Figure 6 shows that both models have the same AUC of 0.82, so they have similar performance in terms of their ability to discriminate between classes.

Additionally, a factor analysis is performed using an ANOVA to understand which factors and combinations of factors significantly affect the dependent variable (LEARNING) and which ones do not have a relevant effect in the study context.

Table 3 shows that Classroom and DEPTH are the only variables that have a significant effect on the dependent variable, with Classroom being highly significant ($p < 0.001$) and DEPTH also significant ($p < 0.05$). This indicates that differences between classrooms and depth have a significant impact on the dependent variable. The rest of the variables and their interactions show no significant effects, suggesting that they do not contribute significantly to the variability observed in the dependent variable. Based on the results of these ANOVA analyses, some hypotheses of the model are supported.

H1: The prompt's effective structure contributes positively to students' depth perception in LLM responses.

Hypothesis H1 has been confirmed. The ANOVA analysis indicates that the variable DEPTH significantly affects LEARNING ($p < 0.05$). This suggests that depth perception influences learning improvement, as seen in **Figure 7**. Qualitative observations from different classrooms further support this finding. Comments such as “the response is solid, but could be improved with more specific details and concrete examples” or “the AI helps better to understand certain aspects, situations, or cases, and can be a good support to expand the basic information” and “the AI provides new perspectives that strengthen our initial argument, but we should not use it as the sole source” reinforce this hypothesis.

Specifically, **Figure 7** shows significant differences in LEARNING variability between the two DEPTH levels. Although the medians are similar, the higher variability at the low DEPTH level could influence LEARNING differently than the high DEPTH level. This information is consistent with the ANOVA results that indicated that DEPTH has a significant effect on LEARNING.

Additionally, **Figure 8** illustrates the significant impact of Classroom variables on learning ($p < 0.001$), indicating variations in learning outcomes across classrooms.

H2: The prompt's effective structure contributes positively to the student's perception of truthfulness in LLM responses.

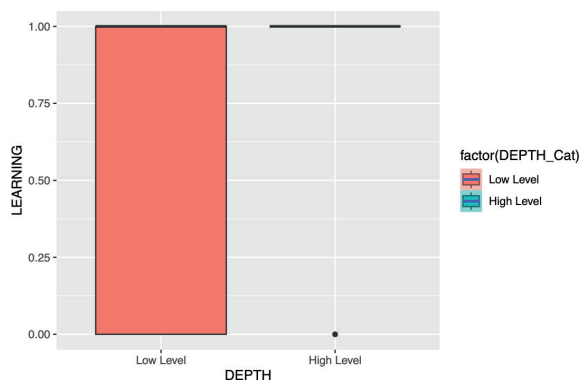


Figure 7.
Boxplot of LEARNING by DEPTH.

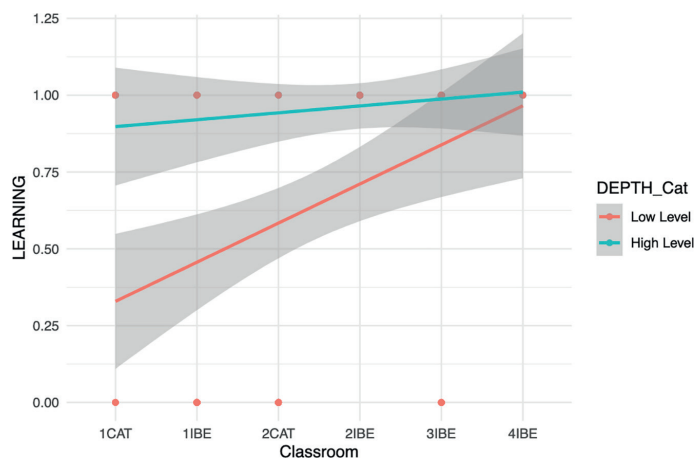


Figure 8.
Interaction between *LEARNING* and *DEPTH* depending on *Classroom*.

The ANOVA results do not allow us to confirm hypothesis H2 due to the elimination of the *VERACITY* variable from the model because of collinearity. Additionally, the evaluation variable (*AI_Mark*), which could be indirectly related to perceived truthfulness, does not show a significant impact on the ANOVA results. Qualitative classroom observations note that “the veracity of the results may also be affected by the quality of the input data and the presence of inherent biases in the data”, suggesting that effective prompt structure and perceived veracity are linked. Qualitative student observations confirm hypothesis H2, indicating that effective prompt structure significantly influences the perceived veracity of responses generated by an LLM. Expressly, in one of the reflections, it is stated, “By using precise language and providing detailed prompts, it increases the chances of receiving accurate responses that match our needs”. The prompt’s precision, specificity, and clarity are determining factors for students to perceive the answers as truthful. Precisely, in another of the reflections, it is mentioned that “depending on the information you give to the AI and the questions you ask, it will guide us to some answers or others. Therefore, when formulating the questions, you must be very

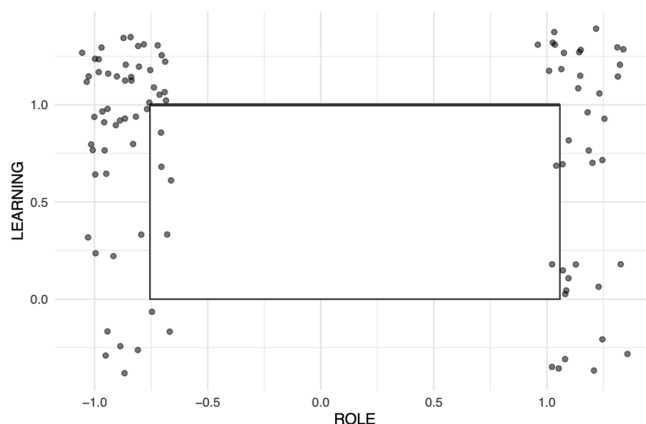


Figure 9.
Distribution of *LEARNING* by *ROLE*.

thorough in providing us with the answers we need and following the conversation thread”, highlighting that the prompt’s quality and specificity affect the response’s quality. This implies that a well-structured prompt improves the perception of truthfulness.

H3: The prompt’s effective structure contributes positively to students’ learning improvement.

The analyses conducted through ANOVA do not support hypothesis H3, as the ROLE does not significantly affect LEARNING (see **Figure 9**). However, it is essential to note that interactions between ROLE and other variables such as Gender, AI_Type, and Classroom could be significant (refer to **Figures 10–12**). **Figure 10** illustrates the variation in LEARNING across different classrooms and roles, indicating that the effectiveness of the prompt can differ significantly depending on the specific classroom and role. **Figure 11** demonstrates the distribution of the ROLE variable across various AI types (AI_Type). Prompt effectiveness generally varies based on the type of AI and the assigned ROLE. **Figure 12** depicts how prompt effectiveness varies

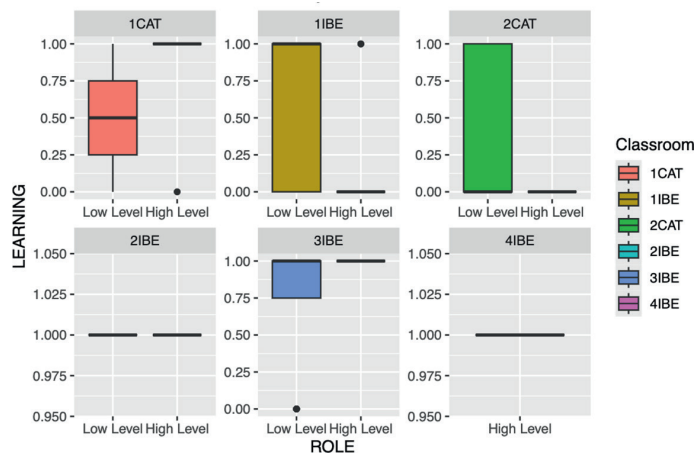


Figure 10.
Distribution of LEARNING by ROLE and Classroom.

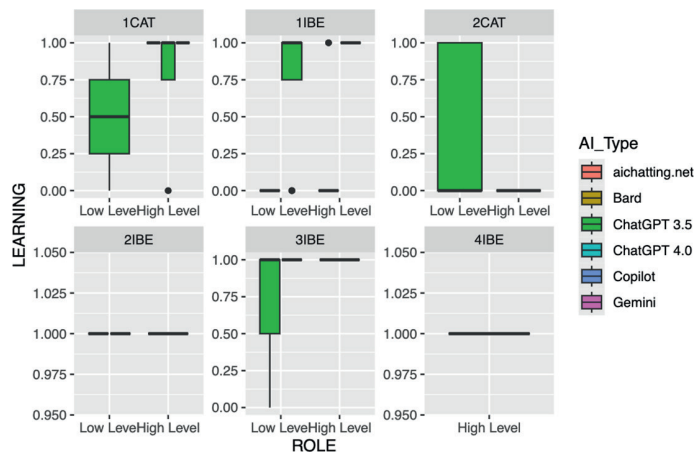


Figure 11.
Distribution of LEARNING by ROLE and AI_Type.

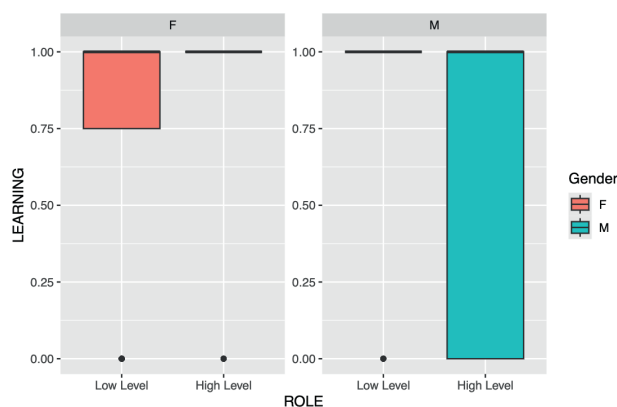


Figure 12.
 Distribution of *LEARNING* by *ROLE* and *Gender*.

by gender and assigned *ROLE*, highlighting notable differences in prompt effectiveness between genders. Classroom reflections reinforce these findings, suggesting that the specificity and clarity of the role in the prompt may influence its perceived effectiveness. One classroom reflection mentioned, “*AI is a very successful tool to help in summaries and generate texts based on very well-worked ideas or prompts, but it lacks reasoning skills*”. This indicates that the role’s specificity and clarity in the prompt may influence the perception of its effectiveness.

The variables *OBJECTIVES*, *CONTEXT*, and *SPECIFIC* have been removed due to lack of variability and collinearity. However, classroom reflections suggest that a clear objective in the prompt is essential. It was mentioned that providing more information to the AI can make the answer more concrete. Additionally, it was suggested that including more specific details and concrete examples can improve the response. Another reflection emphasized the importance of context provided in the prompt, stating that the quality of the AI response depends on the context provided by the user. Although the ANOVA results did not show variability in these variables, hypothesis H3 can be confirmed as there is an importance of including a clear objective,

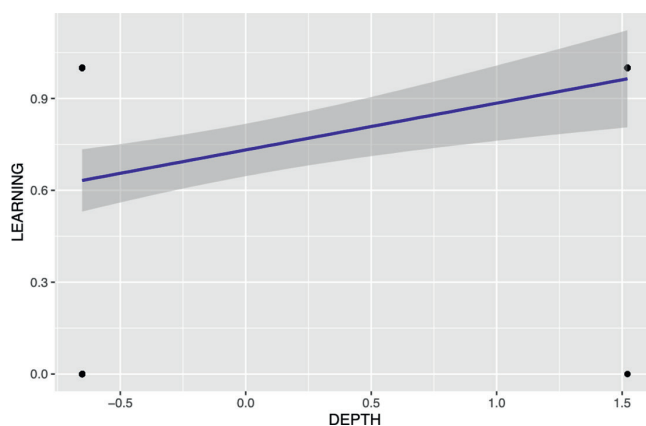


Figure 13.
LEARNING vs. *DEPTH* scatter plot.

contextualizing the information in the prompt, and specifying what is desired as a result of the question asked to the LLMs.

H4: Student's perception of depth contributes positively to learning improvement.

The hypothesis H4 has been confirmed. The depth of AI responses significantly affects learning, implying that depth perception is essential in improving learning (see **Figure 13**). Qualitative student observations further support hypothesis H4, indicating that depth perception in AI responses significantly influences learning enhancement. For instance, one student's reflection indicates that “*AI contributes to reflecting on critical points by correctly identifying areas for improvement*”, suggesting that the adequate depth and structure of the prompt are crucial for learning. Another noteworthy reflection mentions that “*AI helps me to structure the points to comment, to identify key points and concepts, but it does not give me the whole answer. The improvement is due to the ease of use, the structuring of texts, and the summary of concepts, which helps me improve in general*”. While AI helps structure and summarize, real learning improvement stems from deeper reflection, partially facilitated by AI, aiding students in enhancing their understanding and learning.

H5: Student's perception of truthfulness contributes positively to learning improvement.

The ANOVA results do not allow us to confirm hypothesis H5 because the variable VERACITY was eliminated from the model due to collinearity. However, the qualitative evidence collected in the classrooms supports this hypothesis, indicating that students' perception of veracity significantly influences their trust in AI responses and their use of AI to enhance their learning process. The evidence suggests that “*Chat GPT 4.0 is a reliable and valuable resource for comparing proposals and facilitating continuous learning in academic settings if used correctly. The quality and reliability of AI responses are crucial for facilitating continuous learning*”. Students also mentioned that while AI is useful for obtaining information, it should not be the sole source and should be validated with additional information before making decisions. This statement highlights the importance of perceived veracity in AI responses for effective learning. With these arguments, we can partially confirm hypothesis H5.

The ANOVA model confirmed hypotheses H1 and H4, emphasizing the significance of prompt structure for depth perception and its influence on learning enhancement. However, due to the lack of data on VERACITY and other key variables (OBJECTIVE, CONTEXT, SPECIFIC, and EFFECTIVENESS), we could not fully evaluate hypotheses H2, H3, and H5. Nonetheless, the considerable variability between classrooms suggests the necessity to tailor prompts to specific contexts to maximize their effectiveness. While the ANOVA results did not support hypothesis H3, subgroup analysis revealed trends that were not apparent in the overall analysis. Specifically, interactions between ROLE and other variables such as Gender, AI_Type, and Classroom showed marked differences in prompt effectiveness between genders and how the ROLE variable is distributed across different types of AI (AI_Type) and classrooms. This analysis underscores the complexity of the factors influencing the perception and effectiveness of the responses generated by an LLM, emphasizing the importance of considering multiple variables and interactions to enhance educational outcomes.

6. Conclusions

The study examined how the structure of prompts used in LLM models affects the perception of depth, accuracy, and effectiveness of the student's

learning process. The results, obtained through both quantitative and qualitative analyses, confirm several crucial hypotheses regarding the relationship between prompt structure and the perception and effectiveness of AI responses in the learning process.

The main findings of the study indicate that:

1. A well-designed prompt structure contributes positively to the perception of depth and accuracy in LLM responses. This highlights the importance of carefully formulating questions to maximize the usefulness of LLM answers.
2. The positive perception of the depth and accuracy of AI responses significantly improves the student's learning process. This emphasizes the value of comprehensive and truthful responses, not only for user satisfaction but also for long-term educational benefits.
3. The effective structure of the prompt and students' perceptions of accuracy and depth are interrelated, with each enhancing the influence of the other on the perception of the effectiveness of LLM responses.

The results of this study demonstrate the intricate relationship between prompt structure, response perception, and the learning process. By better understanding these dynamics, we can develop strategies to improve prompt design in educational contexts. This will allow us to obtain better responses generated by LLM, ensuring that both prompt effectiveness and students' positive perception of learning are maximized.

However, it is essential to note that the partial significance of the statistical results found in the study has several limitations. Firstly, there was a lack of variability in responses regarding the efficient structure of the prompt and truthfulness of the LLM response, which limited the quantitative analysis and made the results dependent on qualitative data and subjective perceptions. Secondly, the specific use of LLMs for a particular activity means that the results cannot be generalized to all LLM application scenarios or educational disciplines. Finally, although the model was good, a larger sample size could help produce more reliable results.

Below is a set of proposed research ideas that aim to address the limitations observed in the study:

- Expand the database by including a larger and more diverse sample of responses to encompass a broader range of variability in perceived truthfulness and other evaluated metrics.
- Observe whether the type of generative AI used to carry out the activity impacts the confirmation of the hypotheses.
- Conduct experimental studies where the characteristics of the prompts are systematically manipulated to directly evaluate their impact on the quality and perception of AI responses.
- Extend the research to include different educational disciplines and types of questions to explore the applicability and effectiveness of LLMs in various educational contexts.

- Develop analytical tools that enable teachers and developers to evaluate and optimize the structure of prompts in real time.

Acknowledgements

We extend our sincere thanks to all the students who generously volunteered their time and efforts to participate in this study.

Author details

Silvia Rodriguez-Donaire^{1,2}

1 Business and Economics Department, Open University of Catalonia (UOC—Universitat Oberta de Catalunya), Barcelona, Spain

2 UPC-BarcelonaTECH, Universitat Politecnica de Catalunya, ESEIAAT, Terrassa, Spain

*Address all correspondence to: srodriguezdon@uoc.edu;
silvia.rodriguez-donaire@upc.edu

IntechOpen

© 2024 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Park D, An G t, Kamyod C, Kim CG. A study on performance improvement of prompt engineering for generative AI with a large language model. *Journal of Web Engineering*. 2024;**22**(8):1187-1206. DOI: 10.13052/jwe1540-9589.2285
- [2] Megahed FM, Chen YJ, Ferris JA, Knoth S, Jones-Farmer LA. How generative AI models such as ChatGPT can be (mis)used in SPC practice, education, and research? An exploratory study. *Quality Engineering*. 2024;**36**(2):287-315. DOI: 10.1080/08982112.2023.2206479
- [3] Bozkurt A. Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education*. 2023;**18**(1):198-204. Available from: <https://files.eric.ed.gov/fulltext/EJ1389644.pdf>
- [4] Liu L. Analyzing the text contents produced by ChatGPT: Prompts, feature-components in responses, and a predictive model. *Journal of Educational Technology Development and Exchange*. 2023;**16**(1):49-70. DOI: 10.18785/jetde.1601.03
- [5] Cotroneo P, Hutson J. Generative AI tools in art education: Exploring prompt engineering and iterative processes for enhanced creativity. *Metaverse*. 2023;**4**(1):14. DOI: 10.54517/m.v4i1.2164
- [6] Walter Y. Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*. 2024;**21**(1):15. DOI: 10.1186/s41239-024-00448-3
- [7] Schorcht S, Buchholtz N, Baumanns L. Prompt the problem – Investigating the mathematics educational quality of AI-supported problem solving by comparing prompt techniques. *Frontiers in Education*. 2024;**9**:1386075. DOI: 10.3389/feduc.2024.1386075
- [8] Ifenthaler D. Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Journal of Educational Technology & Society*. 2012;**15**:38-52. Available from: <https://www.semanticscholar.org/paper/Determining-the-effectiveness-of-prompts-for-in-Ifenthaler/55065876334df2698da179898d2f1be7501beca1#paper-topics>
- [9] Lo LS. The CLEAR path: A framework for enhancing information literacy through prompt engineering. *Journal of Academic of Librarianship*. 2023;**49**(4):102720. DOI: 10.1016/j.acalib.2023.102720
- [10] Santos J, Figueiredo AS, Vieira M. Innovative pedagogical practices in higher education: An integrative literature review. *Nurse Education Today*. 2019;**72**:12-17. DOI: 10.1016/j.acalib.2023.102720
- [11] Falloon G. From digital literacy to digital competence: The teacher digital competency (TDC) framework. *Educational Technology Research and Development*. 2020;**68**(5):2449-2472. DOI: 10.1007/s11423-020-09767-4
- [12] BaiDoo-Anu D, Owusu Ansah L. Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*. 2023;**7**(1):52-62. DOI: 10.61969/jai.1337500

- [13] Chan CKY, Hu W. Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*. 2023;**20**(1):43. DOI: 10.1186/s41239-023-00411-8
- [14] Maphoto KB, Sevnarayan K, Mohale NE, Suliman Z, Ntsopi TJ, Mokoena D. Advancing students' academic excellence in distance education: Exploring the potential of generative AI integration to improve academic writing skills. *Open Praxis*. 2024;**16**(2):142-159. DOI: 10.55982/openpraxis.16.2.649
- [15] Michel-Villarreal R, Vilalta-Perdomo E, Salinas-Navarro DE, Thierry-Aguilera R, Gerardou FS. Generative AI for higher education as explained by ChatGPT. *Education Sciences*. 2023;**13**(9):856. DOI: 10.3390/educsci13090856
- [16] Ooi KB, Tan GWH, Al-Emran M, Al-Sharafi MA, Capatina A, Chakraborty A, et al. The potential of generative artificial intelligence across disciplines: Prospects and future directions. *Journal of Computer Information Systems*. 2023:1-32. DOI: 10.1080/08874417.2023.2261010 [Ahead of print]
- [17] Pesovski I, Santos R, Henriques R, Trajkovik V. Generative AI for customizable learning experiences. *Sustainability*. 2024;**16**(7):3034. DOI: 10.3390/su16073034
- [18] Ruiz-Rojas LI, Acosta-Vargas P, De-Moreta-Llovet J, Gonzalez-Rodriguez M. Empowering education with generative artificial intelligence tools: Approach with an instructional design matrix. *Sustainability*. 2023;**15**(15):11524. DOI: 10.3390/su151511524
- [19] OpenAI. ChatGPT (Mar 14 Version) [Large Language Model]. 2023. Available from: <https://chat.openai.com/chat> [Accessed: July 17, 2024]
- [20] Eke DO. ChatGPT and the rise of generative AI: Threat to academic integrity? *Journal of Responsible Technology*. 2023;**13**:100060. DOI: 10.1016/j.jrt.2023.100060
- [21] Ocaña-Fernández Y, Valenzuela-Fernández LA, Garro-Aburto LL. Artificial intelligence and its implications in higher education. *Purposes and Representations*. 2019;**7**(2):553-568. DOI: 10.20511/pyr2019.v7n2.274
- [22] Samaniego JF. The four great ethical challenges of applying artificial intelligence to online education. The blog of the UOC's computer science. In: *Multimedia and Telecommunications Studies*. 2022. Available from: <https://blogs.uoc.edu/informatica/es/cuatro-retos-eticos-inteligencia-artificial-educacion-online/> [Accessed: July 17, 2024]
- [23] UNESCO. Artificial Intelligence in Education. Digital Learning and Education Transformation. 2024. Available from: <https://www.unesco.org/es/digital-education/artificial-intelligence> [Accessed: July 17, 2024]
- [24] Dang H, Mecke L, Lehmann F, Goller S, Buschek D. How to Prompt? Opportunities and Challenges of Zero- and Few-Shot Learning for Human-AI Interaction in Creative Applications of Generative Models. (arXiv:2209.01390). arXiv. 2022. Available from: <http://arxiv.org/abs/2209.01390> [Accessed: July 17, 2024]
- [25] Lin Z. How to write effective prompts for large language models. *Nature*

Human Behaviour. 2024;8(4):611-615.
DOI: 10.1038/s41562-024-01847-2

[26] Rodriguez Donaire S. Engineering prompts for codifying students' prompt structure and understanding their learning perception from receiving feedback on an online activity using AI. CORA, Research Data Repository. 2024;V1:UNF:6:VIaUGgwuP9lPi3l4Y BwLZw== [fileUNF]. DOI: 10.34810/data1554