Is ChatGPT helpful for graduate students in acquiring knowledge about digital storytelling and reducing their cognitive load? An experiment.

¿ChatGPT es útil para que los estudiantes de posgrado adquieran conocimientos sobre narración digital y reduzcan su carga cognitiva? Un experimento.

Raidell Avello-Martínez
Universidad Bolivariana del Ecuador, Guayaquil, Ecuador
ravellom@ube.edu.ec

Tomasz Gajderowicz
Universidad de Varsovia, Varsovia, Polonia
tgajderowicz@wne.uw.edu.pl

Víctor G. Gómez-Rodríguez
Universidad Bolivariana del Ecuador, Guayaquil, Ecuador
vgomez@ube.edu.ec

Abstract

This study examines the impact of ChatGPT on narrative scriptwriting abilities and cognitive load in a sample of 41 master's students enrolled in a Digital Narratives course. Using a randomized experimental design, participants were divided into two groups: an experimental group (n = 20) that interacted with ChatGPT and a control group (n = 21) that did not. Our methods involved pre- and post-tests to assess changes in digital storytelling skills and cognitive load, as defined by intrinsic, extraneous, and germane load measures. The results indicated no significant improvement in digital storytelling skills for the experimental group compared to the control group, suggesting that the use of ChatGPT does not markedly enhance narrative writing abilities in the short term. However, a significant reduction in germane cognitive load was observed among the experimental group, pointing to ChatGPT's potential to facilitate the learning process by reducing the mental effort required for task integration and application. The study underscores the complexity of integrating AI into learning environments and highlights the need for strategic AI implementation tailored to specific educational objectives. It also points to the importance of longitudinal research to fully understand the long-term effects of AI on learning and cognitive development.

Keywords: Artificial Intelligence in Education, Digital Storytelling, Cognitive Load Theory, ChatGPT, Narrative Writing Skills.

Resumen

Este estudio examina el impacto de ChatGPT en las habilidades de escritura de guiones narrativos y la carga cognitiva en una muestra de 41 estudiantes de maestría matriculados en un curso de Narrativas Digitales. Utilizando un diseño experimental aleatorio, los participantes se dividieron en dos grupos: un grupo experimental (n = 20) que interactuaba con ChatGPT y un grupo de control (n = 21) que no lo hacía. Nuestros métodos incluyeron pruebas previas y posteriores para evaluar los cambios en las habilidades de narración digital y la carga
Is ChatGPT helpful for graduate students in acquiring knowledge about digital storytelling and reducing their cognitive load? An experiment. Raidell Avello-Martínez, Tomasz Gajderowicz, Víctor G. Gómez-Rodríguez.
Is ChatGPT helpful for graduate students in acquiring knowledge about digital storytelling and reducing their cognitive load? An experiment. Raidell Avello-Martínez, Tomasz Gajderowicz, Víctor G. Gómez-Rodríguez.
contexts. By exploring this issue, it aspires to contribute to the body of knowledge on the application of AI tools in education, addressing both the possibilities these technologies offer to enrich learning and teaching and the challenges their integration presents. Research on the interaction between students and AI technologies is fundamental to understanding not only the potential of these tools to enhance the educational process but also to identify and address the potential risks and limitations that their use may entail.

**Hypothesis development**

The literature offers varied views on ChatGPT’s role in enhancing writing, pointing to the need for a quantitative approach. Accordingly, this study proposes hypotheses to explore ChatGPT’s effect on digital storytelling, aiming for definitive insights:

- **H1** (Baseline Hypothesis): There is no significant difference in digital storytelling knowledge between the experimental and control groups at baseline, confirming initial group equivalence.
- **H2**: Participants will show a significant improvement in digital storytelling scores from pre-test to post-test within each group, indicating ChatGPT’s educational utility.
- **H3**: There will be a significant difference in digital storytelling abilities, specifically in character development, plot structuring, and dialogue crafting, between the experimental and control groups, underscoring ChatGPT’s influence on creative writing components.
- **H4**: Cognitive load differences between groups post-intervention will reveal ChatGPT’s effectiveness in reducing the mental effort required in the storytelling process.

2. **Methodology**

**Study Design**

This study adopted an experimental design (pre-test, post-test) to investigate the impact of using ChatGPT on the acquisition of knowledge about digital storytelling and the reduction of cognitive load in master’s students. A total of 41 students enrolled in the online Digital Narratives course were divided into two groups: an experimental group and a control group, using Google Meet’s random group creation feature to ensure random assignment.

**Participants**

The study involved 41 master’s students enrolled in the Digital Narratives course. Participants were randomly assigned to one of two groups: the experimental group using ChatGPT (n = 20) and the control group (n = 21).

**Procedure**

The study was conducted in three main phases: pre-learning activities, learning phase, and post-intervention assessment.
**Pre-learning activities (10 minutes):** All students completed the following activities:

- Signing of informed consent.
- A prior knowledge test consisting of 3 questions to assess the initial level of knowledge about digital storytelling, using Moodle questionnaire.

**Learning phase (90 minutes):**

- **Group 1 (Experimental):** This group used ChatGPT as a personalized learning support tool. They were assigned three exercises aimed at constructing three essential aspects of digital storytelling scripts: character creation, plot structure, and dialogue elaboration, spending 30 minutes on each exercise.

- **Group 2 (Control):** Students in this group used traditional lesson resources for learning. Like the experimental group, they completed three exercises focused on the same three aspects of storytelling, with 30 minutes allocated for each one.
Is ChatGPT helpful for graduate students in acquiring knowledge about digital storytelling and reducing their cognitive load? An experiment. Raidell Avello-Martinez, Tomasz Gajderowicz, Víctor G. Gómez-Rodríguez.

Figure 1.

Exercise example

<table>
<thead>
<tr>
<th>Group 1 [ChatGPT]</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exercise 1: Character Development (30 minutes)</strong></td>
<td><strong>Exercise 1: Character Development (30 minutes)</strong></td>
</tr>
<tr>
<td><strong>General instructions:</strong></td>
<td><strong>General instructions:</strong></td>
</tr>
<tr>
<td>Describe an initial situation for your story involving the main characters. You can be as detailed as you want. From this situation, you will develop the characters.</td>
<td>Describe an initial situation for your story involving the main characters. Use materials such as articles, YouTube videos, or teacher notes to inspire your character creation.</td>
</tr>
<tr>
<td><strong>Tasks:</strong></td>
<td><strong>Tasks:</strong></td>
</tr>
<tr>
<td>• Open ChatGPT and provide details about the initial situation.</td>
<td>• Research and use various resources to obtain information about the initial situation you have raised.</td>
</tr>
<tr>
<td>• Use the answers from ChatGPT to create a first version of two main characters and two secondary characters that fit this situation.</td>
<td>• Develop two main characters and two secondary characters based on the research carried out.</td>
</tr>
<tr>
<td>• Develop additional questions about the characters, based on ChatGPT answers, to delve deeper into their motivations, conflicts, and relationships.</td>
<td>• Details aspects such as the personality, motivations, conflicts and relationships of the characters.</td>
</tr>
<tr>
<td>• Modify and improve characters according to your preferences and repeat the process with ChatGPT for additional suggestions.</td>
<td>• Adjust and improve characters according to your preferences and the conclusions of your research.</td>
</tr>
</tbody>
</table>

**Post-test assessment (20 minutes):**

- **Cognitive load questionnaire:** A questionnaire was administered to measure the cognitive load experienced by students, based on the work of Klepsch et al. (2017). This instrument measures the dimensions of intrinsic, extraneous, and germane cognitive load. This questionnaire was implemented in Google Forms.

- **Knowledge test:** A post-intervention knowledge test consisting of 6 questions was conducted to assess knowledge acquisition about digital storytelling following the intervention, using Moodle questionnaire (an example in figure 2)

Figure 2

Test question example

Translation

1. Which of the following statements is true about character development?
   a) A character’s motivations and conflicts do not influence the plot.
   b) Character development is not important in a story.
   c) A character’s motivations and conflicts impact the plot.
Measurement Instruments

- **Prior and post-intervention knowledge tests:** Designed by the researcher (teacher of the course) to assess the students’ level of understanding and knowledge about digital storytelling (focus on character creation, plot structure, and dialogue elaboration) before and after the educational intervention. The type of question was selection, with 3 answer options for each question (see previous section).

- **Cognitive load questionnaire:** Based on the methodology developed by Klepsch et al. (2017), this instrument evaluates the different dimensions of cognitive load (intrinsic, extraneous, and germane) experienced by students during the intervention. The Cronbach's alpha value obtained was approximately 0.749, indicating good internal consistency among the items on the scale. This value suggests that the items are well-correlated and collectively contribute to the construct being measured. A Cronbach's alpha value above 0.7 is generally considered acceptable in social and psychological research, implying that the scale is reliable for measuring the intended construct.

Data Analysis

The data analysis section of this study is focused on examining the differences in digital storytelling knowledge and cognitive load between the experimental group (which used ChatGPT) and the control group (which did not use ChatGPT). To this end, two main types of comparisons were conducted:

1. **Group differences in pre-test.** This comparison aimed to identify any differences between the groups in their pre-test (baseline) (H1).

2. **Intra-group paired differences:** This analysis compared the pre-test and post-test results within each group to determine the extent of change in knowledge about digital storytelling and cognitive load as a result of the intervention (H2).

3. **Group differences in post-test:** This comparison aimed to identify any differences between the groups in the post-test (after intervention) scores (H3).

4. **Group differences in cognitive load.** Additionally, for the post-test, cognitive load between the two groups was also compared to assess the impact of ChatGPT's assistance on easing the mental effort required for digital storytelling (H4).

The statistical analysis was conducted using Jamovi (The jamovi project, 2024), focusing on comparing changes in knowledge levels about digital storytelling and cognitive load between the experimental and control groups. To determine the significance of the observed differences in the results of the knowledge tests and cognitive load questionnaires between both groups, appropriate statistical tests were employed.

Below is a table (table 1) summarizing the results of the Levene's test for equality of variances and tests of normality for both pre-test and post-test scores:
The Levene's test results indicate that the variances between the groups are equal for both the pre-test and post-test, as suggested by the non-significant p-values (p > 0.05), indicating homogeneity of variances. This is crucial for the validity of subsequent analyses that assume equal variances across groups.

The tests of normality, conducted using the Shapiro-Wilk test, reveal that the data distribution deviates from normality for both pre-test and post-test scores, as indicated by the significant p-values (p < .001). Because of the deviation from normality we are opting for non-parametric methods for further analysis of the data, given the non-normal distribution of scores.

3. Results

3.1. Demographic analysis

In the comprehensive demographic analysis section of our study, we analyzed the characteristics of the 41 participants. The average age of participants in experimental group (EGp) was 36.1 years, while control group (CGp) had an average age of 32.7 years, leading to an overall average age of 34.4 years for the total sample. The analysis, detailed in Table 2, provides insights into the gender composition of our sample, our findings reveal a balanced representation within the sample. The EGp comprised 8 males and 12 females, while the CGp consisted of 7 males and 14 females, showcasing a near-equitable gender ratio within the study population.

Table 2.
Gender analysis of total sample respondents.

<table>
<thead>
<tr>
<th>Gender</th>
<th>EGp</th>
<th>CGp</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>8</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>Female</td>
<td>12</td>
<td>14</td>
<td>20</td>
</tr>
</tbody>
</table>

Source: Authors’ work
3.2. Pre-test results (H1)

The pre-test, conducted to evaluate baseline conditions between the experimental and control groups, demonstrated minimal initial differences ($p = 0.544 > 0.05$). This outcome, as presented in Table 3, confirms the baseline hypothesis (H1) that both groups possess comparable digital storytelling knowledge before the intervention. Such a uniform starting point ensures that any post-test differences can be attributed to the impact of ChatGPT, establishing a solid base for further analysis.

### Table 3.

Bias determination between samples through pre-test.

<table>
<thead>
<tr>
<th>Group</th>
<th>M</th>
<th>SD</th>
<th>Mean difference</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grupo 1</td>
<td>0.783</td>
<td>0.248</td>
<td>0.053</td>
<td>0.544</td>
</tr>
<tr>
<td>Grupo 2</td>
<td>0.730</td>
<td>0.271</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors’ work

3.3. Intra-group Paired Differences (H2)

The intra-group analysis (Table 4), aimed at assessing H2’s premise that ChatGPT would significantly enhance digital storytelling skills, shows both experimental and control groups improved their scores. However, the statistical analysis, with $p$-values exceeding the 0.05 threshold for both groups, leads to the acceptance of H0 for this hypothesis. This outcome suggests that, despite observed improvements in digital storytelling abilities, the differences induced by ChatGPT’s use are not statistically significant. Thus, the hypothesis anticipating a marked benefit from ChatGPT in knowledge and skill development in digital storytelling was not supported by the data.

### Table 4.

Pre-test and post-test scores of the experimental group and control group.

<table>
<thead>
<tr>
<th></th>
<th>Experimental Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-test</strong></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Mean</td>
<td>0.783</td>
<td>0.248</td>
</tr>
</tbody>
</table>

| **Post-test**                     | Mean | SD  | Mean | SD  | Mean difference | t    | $p$  | Effect Size | Decision |
| Mean                              | 0.900| 0.166| 0.817| 0.166| -0.0873         | -1.39| 0.178| -0.304     | Accept H0 |

Source: Authors’ work

3.4. Group differences in post-test (H3)

The post-test comparison (Table 5), aimed at evaluating differences between groups in digital storytelling ability after intervention (H3), revealed a mean score of 0.900 for Group 1 and 0.817 for Group 2, with a mean difference of 0.0825. Despite the observed difference, the $p$-value of 0.060 suggests that this difference does not reach statistical significance ($p > 0.05$), according to the Mann-Whitney U test. Therefore, while Group 1, which received ChatGPT intervention, showed a higher mean score, this improvement over Group 2, which did not, is
not statistically significant, leading to the acceptance of H0 for H3. This indicates that the impact of ChatGPT on post-intervention digital storytelling skills, while potentially positive, does not demonstrate a clear statistical advantage over traditional methods based on this data set.

Table 5.
Bias determination between samples through post-test.

<table>
<thead>
<tr>
<th>Group</th>
<th>M</th>
<th>SD</th>
<th>Mean difference</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grupo 1</td>
<td>0.900</td>
<td>0.166</td>
<td>0.0825</td>
<td>0.060</td>
</tr>
<tr>
<td>Grupo 2</td>
<td>0.817</td>
<td>0.166</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors’ work Mann-Whitney U

3.5. Group differences in cognitive load (H4)

The analysis showed in figure 3, considering the three types of cognitive load—namely intrinsic (ICL), extraneous (ECL), and germane cognitive load (GCL)—reveals distinct differences between the experimental group (EGp) and the control group (CGp) following the intervention.

- **Intrinsic Cognitive Load (ICL)** reflects the inherent difficulty of the task. The mean scores slightly increased from the control to the experimental group (4.833 to 4.875), indicating a marginal increase in the perceived complexity and demands of the task. However, the difference (0.042) is minimal, suggesting that both groups perceived the task's inherent difficulty similarly.

- **Extraneous Cognitive Load (ECL)**, associated with the way information or tasks are presented to learners, showed a more noticeable difference. The mean scores increased from 5.000 in the control group to 5.233 in the experimental group, with a difference of 0.233. This suggests that the experimental group, potentially influenced by the intervention, experienced a slightly higher extraneous cognitive load. This could imply that the intervention's design or implementation introduced additional complexity or effort in processing the information.

- **Germane Cognitive Load (GCL)**, which pertains to the mental resources allocated to processing and understanding the task, decreased from the control group to the experimental group (4.222 to 3.717), with a difference of -0.506. This significant decrease suggests that the experimental group found it less taxing to integrate and apply the important information, indicating that the intervention may have effectively facilitated learning by reducing the effort needed to understand and link crucial concepts.
3.6. Student perceptions (group interview)

The summary of the perceptions of the students in the experimental group about the activities carried out in the experiment with ChatGPT reveals a predominantly positive reception. Participants found the activities interesting, enriching and entertaining, highlighting the interactivity and novelty of using an artificial intelligence tool of this type for the first time. They expressed that they gained practical benefits and appreciated the help that ChatGPT provided on assignments, finding it especially useful for narrative construction and conflict redesign in their writing.

Several students expressed that the activity was an introduction to a digital tool that they found easy to use and applicable to their educational environments. They appreciated how ChatGPT could save time and help the teaching process, as long as they gave specific prompts to generate useful results. The dynamic nature of building scenes and narratives with the help of AI was also seen as a valuable exercise in creativity and knowledge generation.

However, some students faced challenges and described the activity as complicated or difficult, especially for those who were new to the terminology and process of narrative writing with AI. Despite these initial difficulties, the non-evaluative nature of the activity allowed for a pressure-free exploration of the tool, which was appreciated by the participants.

4. Discussion

Demographic Insights and Baseline Equivalence

This study achieves both gender parity and age diversity, thus ensuring a representative sample that reflects the inherent inclusivity of digital storytelling and AI in education. This inclusion is critical to understanding the multifaceted effects of technology on diverse demographic groups. Avello-Martínez et al. (2023) highlight the potential of digital narratives to cultivate inclusive educational environments, indicating that our well-balanced demographics strengthen the generalizability of our findings.
Pre-Test Results and Baseline Knowledge

Establishing baseline equivalence was an important step in our methodology, ensuring that any post-intervention differences could be confidently attributed to the use of ChatGPT. This approach aligns with the methodological rigor required in educational research, as highlighted by (Essel et al., 2024). The minimal initial differences observed reinforce the importance of carefully controlled experimental designs to evaluate the impact of innovative educational technologies. The result, presented in Table 3, confirmed that both groups had similar knowledge about digital storytelling before the intervention. This starting point established a solid foundation for subsequent analyses.

Intra-Group and Inter-Group Analysis

The observed lack of statistically significant improvement in digital storytelling skills, as determined by intra-group analysis, underscores the complexity of AI's educational impact, particularly regarding ChatGPT. While both experimental and control groups showed improvements, the absence of significant differences underscores the subtlety of AI's role in enhancing educational outcomes. This nuance is echoed in the works of García-Peñalvo (2023) and Muñoz et al. (2024), who argue that AI's effectiveness hinges on thoughtful design and implementation, emphasizing the need for AI to complement rather than replicate traditional pedagogical methods. The call for employing qualitative research methodologies, as Jafari & Keykha (2023) suggest, becomes imperative to capture the depth of students' interactions with AI technologies and their resultant learning experiences. Moreover, the potential for both positive and negative long-term effects of AI on education, highlighted by Pellás (2023), stresses the importance of conducting longitudinal studies. These studies should aim to unravel the enduring impacts of AI on learners' competencies, knowledge acquisition, and attitudes towards learning, offering insights into AI's transformative potential in education. The results from our study, indicating non-significant differences post-ChatGPT intervention, contribute to this ongoing discourse, suggesting that while immediate gains might be subtle, the broader implications and long-term effects of AI in education warrant comprehensive and continued exploration.

Cognitive Load Analysis

Our findings clarify the nuanced interplay between the integration of AI in educational contexts and the management of cognitive load, contributing significantly to the discourse on the effectiveness of digital learning. The subtle increase in intrinsic and extraneous cognitive loads observed following ChatGPT integration underscores the critical need for meticulous instructional design that aligns with human cognitive architecture. This alignment is critical to ensure that the introduction of AI tools into the learning environment does not inadvertently raise cognitive demands to counterproductive levels. In contrast, the notable reduction in relevant cognitive load in the experimental group highlights the potential of AI, such as ChatGPT, to facilitate more efficient learning processes by assisting in the synthesis and application of new knowledge. These findings resonate and expand with the work of Ji et al. (2024), who call for a reevaluation of pedagogical strategies in the digital age, advocating for a balanced approach that leverages the strengths of AI while mitigating its challenges.

5. Conclusions

Our investigation into the effects of ChatGPT on narrative scriptwriting and cognitive load among master's students in educational technology has provided nuanced insights into the
integration of AI in educational settings. Despite achieving gender parity and age diversity, ensuring a representative and inclusive sample, our findings indicate that ChatGPT’s impact on enhancing digital storytelling skills is not statistically significant. This outcome suggests a nuanced role of AI in education, emphasizing the importance of design and implementation in leveraging AI tools effectively. The baseline equivalence established between the experimental and control groups allowed for a controlled assessment of ChatGPT’s impact, underscoring the methodological rigor essential for evaluating innovative educational technologies.

The intra-group and inter-group analyses further revealed the complexities involved in integrating AI into learning environments. Although improvements were observed in both experimental and control groups, the lack of significant differences post-intervention with ChatGPT highlights the subtlety of AI's role in educational outcomes. This finding aligns with the broader discourse on the need for AI to complement rather than replicate traditional pedagogical methods. Additionally, our cognitive load analysis indicates that while there is a slight increase in intrinsic and extraneous cognitive loads, the significant reduction in germane cognitive load suggests ChatGPT's potential to facilitate more efficient learning processes by aiding in the synthesis and application of new knowledge.

Implications and Future Directions

The contributions of this study to the field of educational technology and AI integration are multifaceted. It highlights the potential of AI tools like ChatGPT to support learning processes, albeit within the context of thoughtful implementation and design. The findings suggest that while immediate gains in digital storytelling skills may not be significant, the reduction in cognitive load offers a promising avenue for enhancing learning efficiency. Future research should focus on longitudinal studies to assess the long-term impact of AI on educational outcomes, exploring how AI tools can be optimized to complement traditional teaching methods and reduce cognitive burdens. Moreover, there is a need for qualitative research to understand better the student experience with AI in learning environments. By continuing to explore these areas, educators and researchers can develop more effective strategies for integrating AI into education, maximizing its potential to enrich learning and teaching while addressing the challenges and limitations of its use.

Presentación del artículo: 30 de enero de 2024
Fecha de aprobación: 7 de abril de 2024
Fecha de publicación: 30 de mayo de 2024

Authors’ statement on the use of LLMs

This article has not used texts from (or generated) from an LLM (ChatGPT or others) for its writing.
Funding

This work has not received any specific grants from funding agencies in the public, commercial, or non-profit sectors.

6. References


The jamovi project. (2024). Jamovi (Version 2.3)[Computer Software]. https://www.jamovi.org/about.html

