

Towards Player-Oriented Procedural Narratives with Stochastic Grammar-Generated Quests

Luiza Torello Vieira¹, Leonardo Tórtoro Pereira², Claudio F. M. Toledo¹

¹ Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo (USP)
São Carlos – SP – Brazil

²São Paulo State University (UNESP)
Institute of Geosciences and Exact Sciences
Rua 24 A – 13506-900 – Rio Claro – São Paulo – Brazil

luiza.torello@usp.br, claudio@icmc.usp.br, leonardo.t.pereira@unesp.br

Abstract. Introduction: Procedural quest generation has the potential to improve replayability and player engagement, but existing approaches often lack adaptation to different player profiles. **Objective:** This work investigates whether a player-adaptive quest generation method can increase content variability and align generated quests with distinct player behavior profiles. **Methodology:** We propose a quest generation system that combines stochastic grammars for structural coherence with Markov-based selection for variability, incorporating player profiling to guide the generation process. The method is evaluated against a random baseline under different parameter configurations. **Results:** The results of exploratory studies indicate increased diversity and reduced repetition compared to random generation, as well as partial evidence of alignment between generated content and player profiles, although this effect depends on parameter tuning. **Keywords** stochastic grammar, quest generation, procedural content generation.

1. Introduction

Procedural narrative generation is widely used in applications such as video games and simulations. In this context, quests—which are composed of tasks and rewards [Yu et al. 2021]—play a key role in guiding players and supporting interactive, personalized experiences [Soares de Lima et al. 2019]. Despite its importance, procedural quest generation has been explored in prior work, but remains insufficiently studied [Doran and Parberry 2011, Liapis 2020], given its strong potential to improve variability, replayability, and player engagement. This paper proposes a method for generating player-oriented quests using a stochastic grammar combined with a Markov Chain. The grammar defines possible quest structures, while the Markov Chain ensures diversity and reduces repetition. We evaluate the approach through a game prototype and both quantitative and qualitative studies, obtaining promising results.

We hypothesize that this method can generate high-quality quests while ensuring variety and adapting to different player profiles. To investigate this, we address the following research questions: (RQ1) how the method compares to random generation; (RQ2) whether it ensures variety; and (RQ3) whether it improves player enjoyment when aligned with player profiles.

The main contributions of this study are: (i) the design of a procedural quest generation method that integrates structured generation techniques with probabilistic selection to balance coherence and variability, (ii) the incorporation of player profiling into the generation process, enabling the creation of quests tailored to different player behavior patterns, (iii) an experimental evaluation comparing the proposed method with baseline random generation, analyzing its impact on quest variety and alignment with player profiles, and (iv) an analysis of the relationship between player-adaptive quest generation and perceived player enjoyment.

2. Related Work

Player profiling has been widely studied through approaches based on observed behavior [Rivera-Villicana et al. 2018] and player preferences [Vahlo et al. 2017, Yee and Ducheneaut 2018]. Early models such as [Bartle 2003] evolved into motivation-driven classifications, including Yee’s motivational framework. In this work, we adopt the MAIC model—Mastery, Achievement, Immersion, and Creativity—obtained through self-assessment methods [Jameson 2002, Pereira et al. 2022], as it emphasizes players’ intrinsic motivations rather than interaction-based categories, making it better suited for generating personalized quest content.

Quests can be understood as structured sub-stories composed of atomic actions [Doran and Parberry 2011, Breault et al. 2018], a representation widely reused in the literature [Alvarez et al. 2021] and adopted in our system. Existing quest generation approaches include planning-based methods [Breault et al. 2018], multi-system storytelling [Montfort et al. 2013], hybrid techniques [Gellel and Sweetser 2020], and player-adaptive systems [Thue et al. 2021, Lee and Cho 2011]. Despite these advances, challenges such as maintaining coherence remain [Alvarez et al. 2021], and surveys highlight the growing interest in player-adaptive generation [Lopes and Bidarra 2011].

Grammar-based approaches are widely used in procedural content generation to ensure structural coherence. For example, [Kaweeratanakit and Kotrajaras 2025] proposes a grammar for generating complex enemy behaviors, while recent work combines grammars with large language models to guide content generation. However, such approaches may face limitations in expressiveness and scalability, particularly when relying on predefined rule sets [Oğuz and Dockhorn 2024].

In parallel, probabilistic methods such as Markov Chains have been successfully applied to model sequential dependencies in content generation, including levels and music [Moore et al. 2018, Ramanto and Maulidevi 2017, Roig et al. 2018]. These approaches enable the generation of varied and coherent outputs, as demonstrated in game level generation pipelines [Robinet 2024], and have also been applied to simulate narrative and behavioral data [Pérez et al. 2024].

Although grammars and Markov-based methods have been widely explored in procedural content generation, their application to quest generation remains limited. To address this gap, our work combines structured grammar-based generation with Markov modeling. Unlike [Pereira et al. 2022], who rely on static grammar constructs, our approach introduces temporal dependencies and probabilistic variability for non-deterministic generation and repetition avoidance. In contrast to [Soares de Lima et al. 2019], whose planning-based method requires full world-state

specifications and explicit goals, our grammar-based method offers greater flexibility by decoupling quest structure from the world state and enabling dynamic length control.

3. Methodology

Player preferences are modeled through a pre-game questionnaire designed to capture four player profiles: *Creativity (C)*, *Achievement (A)*, *Mastery (M)*, and *Immersion (I)*. This questionnaire is the same as used by [Pereira et al. 2022]. Each question contributes to one or more profiles with associated weights, producing a score ps_i for each profile. These scores are normalized to define the probability of selecting each profile during quest generation. The normalization in equation (1) results in a probability distribution over player profiles, guiding the generation process. Profiles with higher scores are more likely to influence the generated quest lines.

$$pref_i = \frac{ps_i}{\sum_{j=1}^4 ps_j} \quad (1)$$

We employ a stochastic grammar $Gr = (N, \Sigma, P, S)$ to generate quest lines. Non-terminal symbols represent player profiles, while terminal symbols correspond to quest types. In Equation 2, we present the grammar used in our prototype.

$$\begin{aligned} N &= \{S, C, A, M, I\} \\ \Sigma &= \{\text{get, spy, damage, take, use, stealth, capture, defend, repair, steal,} \\ &\quad \text{experiment, escort, explore, goto, gather, exchange, kill, listen, read, give, report, } \varepsilon\} \\ S &\rightarrow C \mid A \mid M \mid I \\ C &\rightarrow \text{escort } C \mid \text{goto } C \mid \text{stealth } C \mid \text{spy } C \mid \varepsilon \\ A &\rightarrow \text{gather } A \mid \text{experiment } A \mid \text{take } A \mid \text{repair } A \mid \varepsilon \\ M &\rightarrow \text{kill } M \mid \text{defend } M \mid \text{capture } M \mid \text{damage } M \mid \varepsilon \\ I &\rightarrow \text{listen } I \mid \text{read } I \mid \text{give } I \mid \text{report } I \mid \text{give } I \mid \varepsilon \end{aligned} \quad (2)$$

Each non-terminal generates quests associated with a specific player profile (e.g., C for exploration and M for combat), while ε denotes the end of a quest line. Quest generation starts from S and iteratively applies production rules until ε is reached. At each step, the next symbol is selected through a Markov chain, whose transition probabilities depend only on the current symbol, enabling coherent yet varied quest sequences¹.

The probability of selecting a non-terminal symbol from S is given by the player profile distribution defined in Equation (1), which ensures that quest generation is personalized according to player preferences. The probability of expanding a non-terminal into terminal symbols or terminating the derivation is controlled by a function.

Let $f_i(x)$ denote the probability of selecting a terminal symbol and $e_i(x)$ the probability of selecting ε , where x is the current length of the quest line. These functions satisfy:

¹Click here to view the Markov probability study.

$$f_i(x) = \frac{1 - e_i(x)}{n} \quad (3)$$

where n is the number of possible terminal symbols for a given profile. The function $e_i(x)$ is defined to increase with x , ensuring that the probability of terminating the quest grows as the quest becomes longer. This formulation allows the system to dynamically balance quest length, avoiding both excessively short and overly long quest lines. The function $e_i(x)$ used in our experiments was as follows:

$$e_i(x) = (0.5x)^2 \quad (4)$$

To validate the proposed approach, we developed a top-down game prototype²³ with basic mechanics, including movement, combat, item collection, and interaction with game agents. The same questionnaire used for quest generation also informs the procedural generation of dungeons and enemies, ensuring consistency across game elements. The implemented quest types include *explore*, *goto*, *gather*, *exchange*, *kill*, *listen*, *read*, *give*, and *report*. These were selected based on feasibility within the prototype and their relevance to the target game genre. Each generated quest includes parameters that define the required gameplay elements. These parameters are essential for integrating the quest system with dungeon and content generation.

4. Results

We evaluated the quest generator using eight different player profile configurations, each executed 200 times, and conducted a quantitative and qualitative evaluation of the results. Since there was no need to implement quests in the prototype, we were able to test the system with all quest type.

4.1. Quantitative Evaluation

Figure 1 presents a representative example of the distribution of generated quests across two runs of our pilot study⁴. The goal was to analyze how input weights influence the distribution of generated quest types.

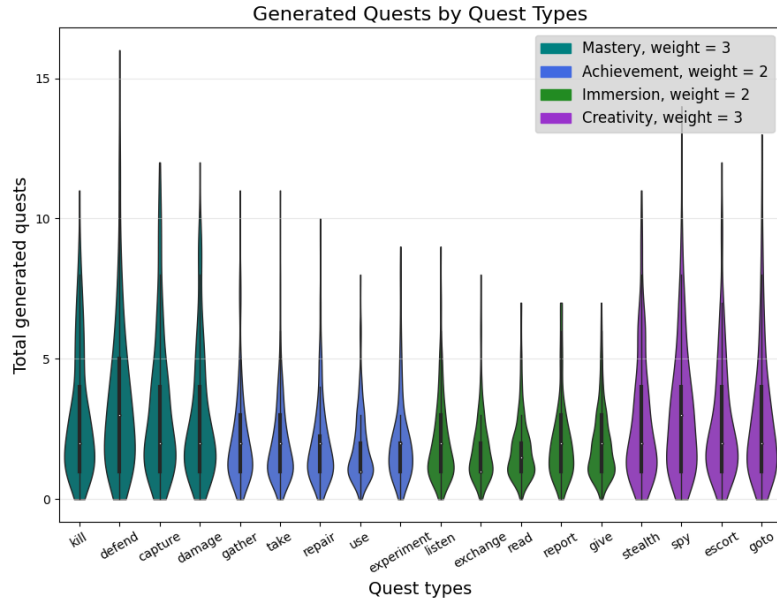
The results show that quest frequencies follow the expected proportions defined by the input profiles. In Figure 1a, we assigned mastery and creativity the same weight (3), and immersion and achievement the weight 2. It is possible to observe that similar violin plots mean the same weight. We also had clearly many outliers which show the stochastic nature of our method. In Figure 1b, where we had different weights for every profile, we can provide evidence that profiles with lower weights consistently produce fewer quests, while higher-weighted profiles dominate the generated quest lines.

To summarize results across all configurations, Figure 2 shows the mean number of generated quests per profile type (x axis) and input profile (y axis). The heatmap highlights a clear relationship between input weights and output distribution: higher profile weights lead to a greater number of corresponding quest types. At the same time,

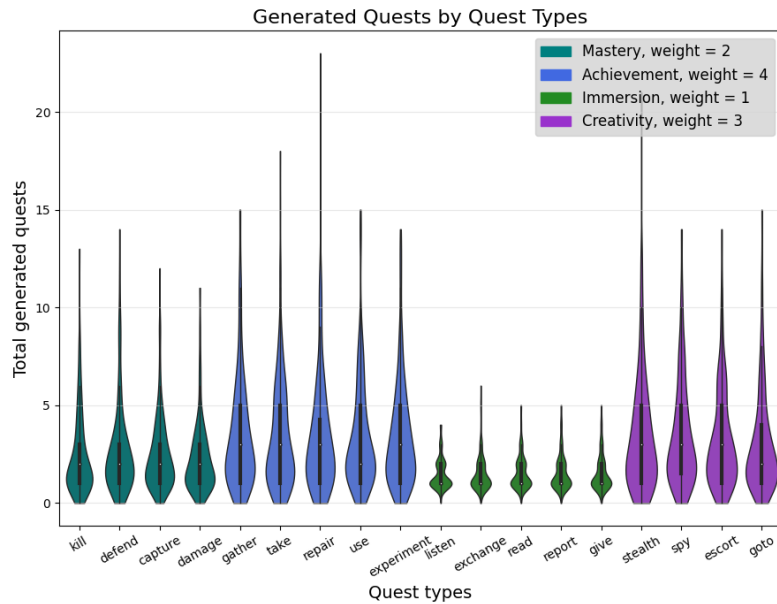
²Click here to view the code for the game prototype

³Click here to download our prototype's latest build

⁴Click here to view the questlines generated in one of those runs for profile M3A2C1I4



(a) Experiment with similar weights.



(b) Experiment with different weights.

Figure 1. Results of quantitative experiments with grammar 1.

variability across runs indicates that the stochastic grammar introduces diversity even under identical inputs. In the first line of Figure 2, we have a pure random method to compare with our own: it has an equal probability to select each non-terminal, then an equal probability for each terminal.

To assess statistical significance, we applied the Kruskal-Wallis test, as normality assumptions were not satisfied. Post-hoc comparisons were performed using Dunn's test. Results indicate that differences between input configurations are statistically significant in most cases ($p < 0.05$), confirming that the generator responds to changes in player profiles. Non-significant differences primarily occur between configurations with

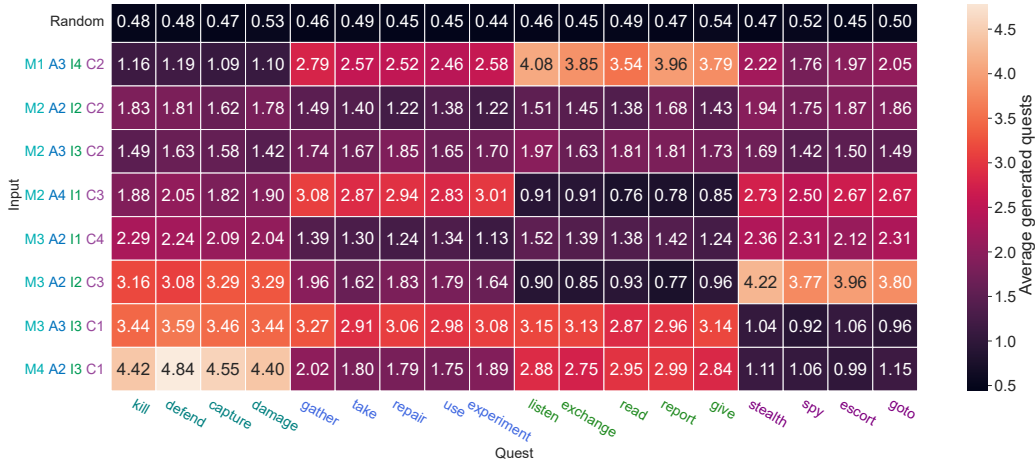


Figure 2. Generated quests of the same profile type for each input averaged from 200 executions with grammar 1.

identical or very similar weights. These findings demonstrate that the proposed method can effectively adapt quest generation according to player preferences while maintaining variability.

Table 1 summarizes which cells could not be rejected as having the same mean between inputs. As expected, most of the paired inputs are the ones with the same weight for the given quests. The exceptions are some inputs with a single unit of difference. The two bottom-most rows show the pairs that have appeared for more than one type of quest, meaning their outputs are more similar than the others. We highlight that the input *M2A2I2C2* was present in both, which may show that providing the same inputs for all profile types can bias the algorithm.

To evaluate repetitiveness, we measured the distance between occurrences of quests of the same type within a quest line. Results seen in Figure 3 show that higher-frequency quest types tend to have shorter distances between occurrences, while lower-frequency types are more spaced out. Importantly, the average distance remains above a minimum threshold in all cases, indicating that the system avoids excessive repetition. We highlight that this Figure inherently shows inverse relationships—higher weights yield more quests. Statistical analysis confirms that these differences are significant in most cases. This behavior is attributed to the Markov-based generation process, which penalizes consecutive repetitions of the same quest type. Overall, the results demonstrate that the method ensures both diversity in quest types and adequate spacing between similar quests.

4.2. Qualitative Evaluation

Since our qualitative study required a playable prototype, we had to remove quest actions that would be too complex to implement and modify the grammar to be as described in Equation 5. The Mastery profile had only one quest. We decided to move Exchange into Achievement and add Explore to Creativity so that the same would not happen to the other profiles.

Table 1. For each quest type, each pair of inputs that could not be rejected as having the same mean of generated quests in grammar 1. Note: — marks insufficient data for more input pairs

Achievement	Creativity	Immersion	Mastery
M2A4I1C3 M1A3I4C2	M2A2I2C2 M1A3I4C2	M3A2I2C3 M2A4I1C3	M2A4I1C3 M2A2I2C2
M3A3I3C1 M1A3I4C2	M3A2I1C4 M2A4I1C3	M4A2I3C1 M3A3I3C1	M3A2I1C4 M2A4I1C3
M4A2I3C1 M3A2I2C3	M3A2I3C1 M3A3I3C1	—	M3A3I3C1 M3A2I2C3
M3A2I2C3 M2A3I3C2	—	—	—
M4A2I3C1 M2A3I3C2	—	—	—
M3A3I3C1 M2A4I1C3	—	—	—
M3A2I1C4 M2A2I2C2	M3A2I1C4 M2A2I2C2	M3A2I1C4 M2A2I2C2	—
—	—	M2A3I3C2 M2A2I2C2	M2A3I3C2 M2A2I2C2

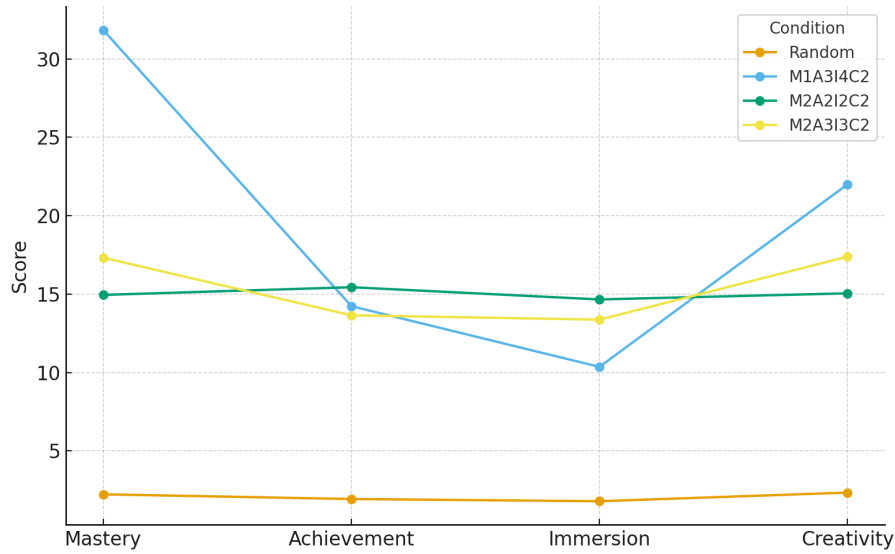


Figure 3. Mean distance of quests of the same profile type for each input averaged from 200 executions with grammar 1.

$$\begin{aligned}
 N &= \{S, C, A, M, I\} \\
 \Sigma &= \{\text{explore, goto, gather, exchange, kill, listen, read, give, report, } \varepsilon\} \\
 S &\rightarrow C \mid A \mid M \mid I \\
 C &\rightarrow \text{explore } C \mid \text{goto } C \mid \varepsilon \\
 A &\rightarrow \text{gather } A \mid \text{exchange } A \mid \varepsilon \\
 M &\rightarrow \text{kill } M \mid \varepsilon \\
 I &\rightarrow \text{listen } I \mid \text{read } I \mid \text{give } I \mid \text{report } I \mid \varepsilon
 \end{aligned} \tag{5}$$

To evaluate player experience, we conducted a user study with 10 participants from our research group. A total of 119 playthroughs⁵⁶⁷ were recorded, with 74 completed runs included in the analysis. Players reported their enjoyment after each dungeon using a 5-point Likert scale ("I had fun completing the quests"). Figure 4 shows the average fun ratings across player profiles and dungeon configurations. To increase the number of samples per group, profile weights were aggregated into low and high preference ranges (multiples of 10). Given the limited sample size, the results should be interpreted as exploratory, as supported by the power analysis in Table 2.

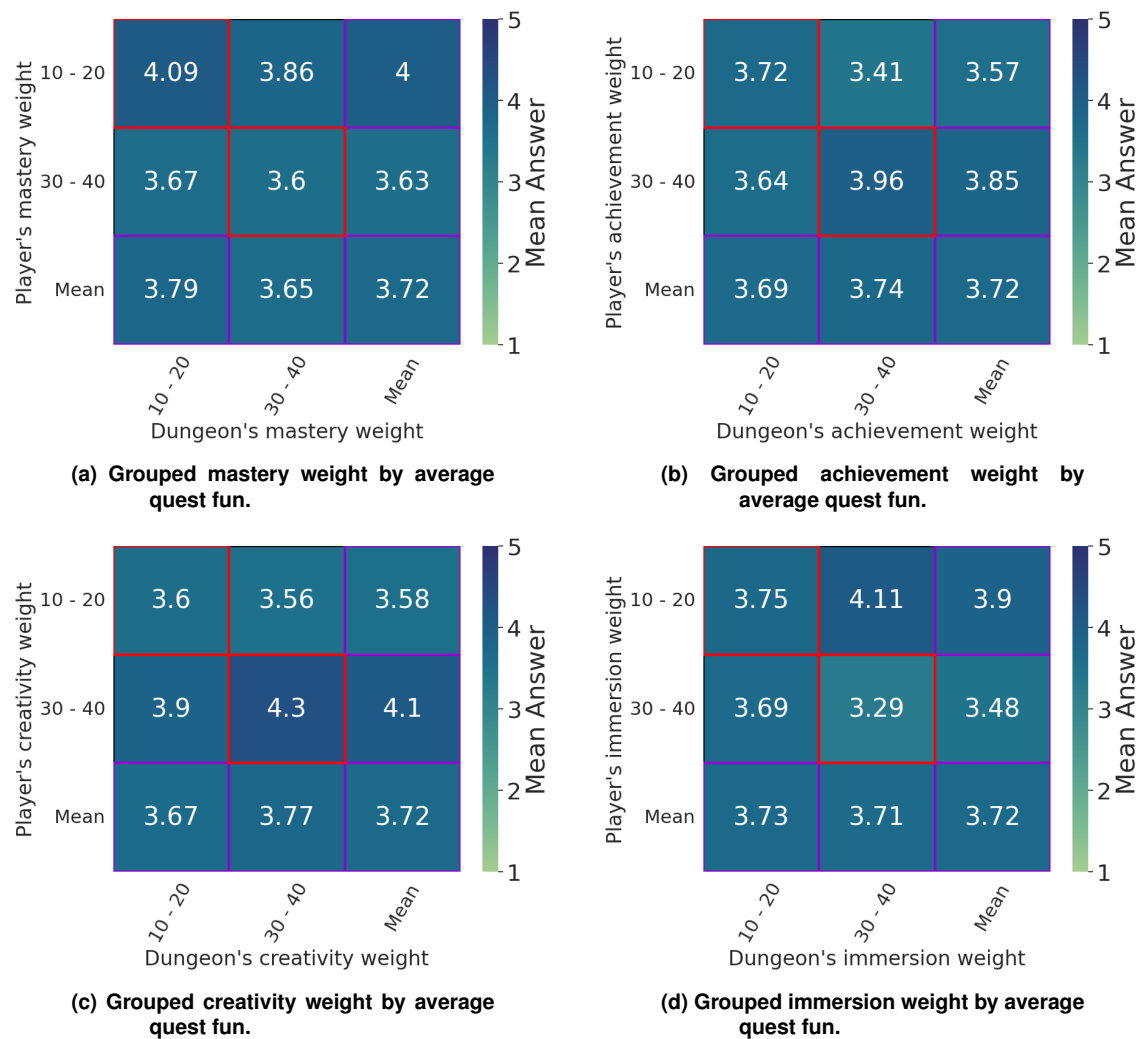


Figure 4. Average answer to the Likert scale quest fun, grouped by player and dungeon profile in grammar 2.

Results indicate a general tendency for higher enjoyment when dungeon content aligns with the player's profile (highlighted in red). In most cases, matching player and dungeon profiles resulted in higher or above-average fun scores compared to non-matching configurations. This suggests that the proposed method can generate content

⁵Click here to view the full user data json.

⁶Click here to view the code used to generate graphs

⁷Click here to view the pre-game questionnaire and player data.

that better fits individual player preferences. However, this trend is not uniform across all profiles. While achievement- and creativity-oriented configurations showed more consistent improvements, immersion-related results were less conclusive. These differences may be related to the specific quest types available in the prototype or to the way information is grouped. Because of the limited number of participants and uneven distribution of samples across configurations, these results should be interpreted as indicative rather than conclusive. Some profile combinations had a small number of observations, which limits statistical validation.

Overall, the qualitative results suggest that the proposed approach can adapt quest generation to different player preferences, thereby increasing perceived enjoyment. These findings complement the quantitative results, indicating that the system produces both measurable and perceptible improvements in generated content.

To further investigate whether profile-aligned content improves the player experience, we conducted statistical hypothesis tests on post-test questionnaire responses. Normality was assessed using the Shapiro-Wilk test, and, depending on the outcome, either a Student's t-test or a Mann-Whitney U test was used. A one-tailed test ($\alpha = 0.05$) was used based on the observed tendencies. Due to the limited number of samples per group, statistical power is sufficient only for large effect sizes. Therefore, results should be interpreted with caution.

Table 2. Comparison of matching player profiles in a generated dungeon with grammar 2. A plus (+) or minus (-) after the p-value indicates positive or negative impact of the algorithm on player experience if the null hypothesis is rejected. Significant results ($p < 0.05$) are highlighted in blue (lesser tail) or red (greater tail). The last two rows show test power for Cohen's $d = 0.8$ (large effect, purple highlight) and $d = 1.2$ (very large effect, yellow highlight); power is acceptable when above 80%.

Profile	Achievement				Mastery			
Weights	P:10-20	P:30-40	D:10-20	D:30-40	P:10-20	P:30-40	D:10-20	D:30-40
Normality p-value	.000	0.096	.000	.008	.001	.001	.001	0.294
Hypothesis p-value	0.076+	0.159+	0.239+	.040+	0.343+	0.301-	0.244+	0.285-
Tail side	GREAT	LESS	GREAT	LESS	GREAT	GREAT	GREAT	GREAT
Power D=0.8	0.749	0.000	0.709	0.000	0.476	.909	0.709	0.589
Power D=1.2	.966	0.000	.950	0.000	0.767	.998	.950	.877
Profile	Immersion				Creativity			
Weights	P:10-20	P:30-40	D:10-20	D:30-40	P:10-20	P:30-40	D:10-20	D:30-40
Normality p-value	.004	.009	.004	.003	.004	.026	.004	.003
Hypothesis p-value	0.107-	0.298-	0.459+	0.061-	0.528+	0.107+	0.236-	.044+
Tail side	LESS	GREAT	GREAT	GREAT	GREAT	LESS	LESS	LESS
Power D=0.8	0.000	0.726	0.785	0.749	.898	0.000	0.000	0.000
Power D=1.2	0.000	.958	.978	.966	.997	0.000	0.000	0.000

Table 2 summarizes the hypothesis testing results. Four statistically significant cases were observed, three of which support a positive effect of profile alignment, with matching configurations receiving higher evaluation scores. One significant negative result was found for immersion, where non-matching configurations performed better, possibly due to the prototype's limited narrative depth and dialogue mechanics.

Overall, the statistical analysis provides partial support for the hypothesis that aligning generated content with player profiles improves player experience. While not all results are conclusive, they reinforce the trends observed in the qualitative analysis.

5. Conclusion

This paper presented a player-adaptive quest generation approach that combines stochastic grammars with Markov-based probability modeling to balance structural coherence and variability. The method enables the generation of personalized quest lines while increasing diversity and reducing repetition. Quantitative results show that the system produces differentiated quest distributions aligned with player profiles, while the Markov-based mechanism contributes to reducing repetitiveness.

Results from the user study suggest that aligning generated content with player profiles can improve perceived enjoyment, although the statistical analysis provides only partial support for this effect. In this context, the findings address the research questions by demonstrating that the proposed method outperforms random generation (RQ1), increases diversity (RQ2), and partially supports player-profile alignment (RQ3), with results dependent on parameter configuration.

This work contributes a player-adaptive quest generation method and an empirical evaluation of its effects on diversity and profile alignment, highlighting the importance of parameter tuning. Despite these contributions, the limited sample size and the simplified implementation of certain player profiles, such as immersion, restrict the generalizability of the results.

While our current baseline demonstrates clear improvements over random generation, we acknowledge that a gold-standard comparison would further strengthen the evaluation. In future work, we intend to conduct a gold-standard evaluation by comparing our generated quests against manually designed quests created by domain experts. Additionally, we intend to investigate the use of higher-order Markov chains to capture longer-range dependencies between quest elements, potentially improving narrative coherence and structural consistency. We also intend to improve narrative depth in the prototype, conduct larger-scale user studies, and extend the approach to other domains such as dynamic difficulty adjustment and educational content generation.

Declaration of Generative AI and AI-assisted technologies

We used LLMs such as DeepSeek and ChatGPT to improve the article's overall quality. We take full responsibility for all content produced with AI assistance.

References

Alvarez, A., Grevillius, E., Olsson, E., and Font, J. (2021). Questgram [qg]: Toward a mixed-initiative quest generation tool. In *Proceedings of the 16th International*

- Conference on the Foundations of Digital Games*, FDG '21, New York, NY, USA. Association for Computing Machinery.
- Bartle, R. (2003). *Designing Virtual Worlds*, pages –768.
- Breault, V., Ouellet, S., and Davies, J. (2018). Let CONAN tell you a story: Procedural quest generation. *CoRR*, abs/1808.06217.
- Doran, J. and Parberry, I. (2011). A prototype quest generator based on a structural analysis of quests from four mmorpgs. In *Proceedings of the 2nd International Workshop on Procedural Content Generation in Games*, PCGames '11, New York, NY, USA. Association for Computing Machinery.
- Gellel, A. and Sweetser, P. (2020). A hybrid approach to procedural generation of roguelike video game levels. In *International Conference on the Foundations of Digital Games*, pages 1–10.
- Jameson, A. (2002). *Adaptive Interfaces and Agents*, page 305–330. L. Erlbaum Associates Inc., USA.
- Kaweeratanakit, N. and Kotrajaras, V. (2025). Towards the pcg-g in shmups: Enemy pattern description language (epdl).
- Lee, Y.-S. and Cho, S.-B. (2011). Context-aware petri net for dynamic procedural content generation in role-playing game. *Computational Intelligence Magazine, IEEE*, 6:16 – 25.
- Liapis, A. (2020). 10 years of the pcg workshop: Past and future trends. In *International Conference on the Foundations of Digital Games*, FDG '20, New York, NY, USA. Association for Computing Machinery.
- Lopes, R. and Bidarra, R. (2011). Adaptivity challenges in games and simulations: A survey. *IEEE Transactions on Computational Intelligence and AI in Games*, 3(2):85–99.
- Montfort, N., y Pérez, R. P., Harrell, D. F., and Campana, A. (2013). Slant: A blackboard system to generate plot, figuration, and narrative discourse aspects of stories. In *ICCC*, pages 168–175.
- Moore, J., Corrêa, D., and Small, M. (2018). Is bach's brain a markov chain? recurrence quantification to assess markov order for short, symbolic, musical compositions. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28:085715.
- Oğuz, M. K. and Dockhorn, A. (2024). Markov senior – learning markov junior grammars to generate user-specified content.
- Pereira, L., Viana, B., and Toledo, C. (2022). A system for orchestrating multiple procedurally generated content for different player profiles. *IEEE Transactions on Games*, 16:64–74.
- Pérez, J., Castro, M., Awad, E., and López, G. (2024). Generation of probabilistic synthetic data for serious games: A case study on cyberbullying. *Knowledge-Based Systems*, 286:111440.

- Ramanto, A. S. and Maulidevi, N. U. (2017). Markov chain based procedural music generator with user chosen mood compatibility. *International Journal of Asia Digital Art and Design Association*, 21(1):19–24.
- Rivera-Villicana, J., Zambetta, F., Harland, J., and Berry, M. (2018). Informing a bdi player model for an interactive narrative. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '18, pages 417–428, New York, NY, USA. ACM.
- Robinet, L. D. (2024). Procedural content generation in celeste overview of automated level generation in video games with a focus on playability in a 2d platformer. Master's thesis, LIACS, Leiden University.
- Roig, C., Tardón, L. J., Barbancho, I., and Barbancho, A. M. (2018). A non-homogeneous beat-based harmony markov model. *Knowledge-Based Systems*, 142:85–94.
- Soares de Lima, E., Feijó, B., and Furtado, A. L. (2019). Procedural generation of quests for games using genetic algorithms and automated planning. In *2019 18th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, pages 144–153.
- Thue, D., Bulitko, V., Spetch, M., and Wasylishen, E. (2021). Interactive storytelling: A player modelling approach. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 3(1):43–48.
- Vahlo, J., Kaakinen, J. K., Holm, S. K., and Koponen, A. (2017). Digital game dynamics preferences and player types. *Journal of Computer-Mediated Communication*, 22(2):88–103.
- Yee, N. and Ducheneaut, N. (2018). *Gamer motivation profiling: Uses and applications*, pages 485–490.
- Yu, K. K., Guzdial, M., and Sturtevant, N. R. (2021). The definition-context-purpose paradigm and other insights from industry professionals about the definition of a quest.