

Mission Profiles and Interaction Patterns in Gravitas: Evidence from a Simulation-based Physics Learning Game

Katharina Richter, Dominik Tschirky and Michael D. Kickmeier-Rust

St. Gallen University of Teacher Education, Switzerland

katharina.richter@phsg.ch

dominik.tschirky@phsg.ch

michael.kickmeier@phsg.ch

Abstract: Digital games and simulation-based environments are recognized as promising approaches for enhancing conceptual understanding and motivation in physics education, particularly with regard to persistent misconceptions in mechanics. *Gravitas* is a simulation-based learning game for lower-secondary students that employs a cyclical design of diagnosis, cognitive conflict, simulation, and reflection to address misconceptions about gravity and forces. In a quasi-experimental field study, 272 students selected one of four self-chosen mission profiles (Adventurer, Researcher, Tactician, Tinkerer). Log data and reflection-based performance measures were analyzed, with profiles serving as a categorical grouping variable. Interaction metrics (simulation time, reset frequency, use of conceptual supports, reflection duration) and reflection scores were normalized per completed level and examined using Kruskal-Wallis tests and Spearman correlations. Results revealed no significant differences between profiles in game progress, exploration intensity, support usage, reflection duration, or reflection performance (all $\epsilon^2 \approx 0$). Reset frequency showed a marginal profile effect ($\chi^2(3) = 7.59$, $p = .055$, $\epsilon^2 = .02$) and a small negative correlation with reflection performance ($\rho = -.18$, $p = .006$). Overall, the findings indicate that self-selected mission profiles had limited explanatory value for learners' interaction patterns and in-game conceptual performance in *Gravitas*. This suggests that adaptive design in simulation-based learning games may benefit from relying more strongly on dynamic, behavior-based indicators rather than on static self-categorizations.

Keywords: Simulation-based learning, Physics education, Conceptual change, Serious games, Learner profiling, Log data analysis

1. Introduction

Digital games and simulation-based environments are widely regarded as promising approaches for addressing both motivational and cognitive challenges in physics education. A substantial body of research shows that interactive simulations and well-designed serious games can foster conceptual understanding and engagement compared to traditional instruction (e.g., Wieman et al., 2008; Shute et al., 2013; Wouters et al., 2013).

In mechanics specifically, persistent misconceptions about gravity, forces, and motion remain a major obstacle despite curricular reforms and conventional teaching. Students often believe, for example, that heavier objects fall faster or that gravity ceases in space; such intuitive conceptions are cognitively stable and frequently survive traditional instruction (Halloun and Hestenes, 1985; McCloskey et al., 1983).

Conceptual change research suggests that overcoming such misconceptions requires more than delivering correct information: learners must experience cognitive conflict, recognize the inadequacy of their prior beliefs, and find alternative conceptions intelligible, plausible, and fruitful (Posner et al., 1982; Limón, 2001). Conceptual change theory, together with Self-Determination Theory (SDT) and Achievement Goal Theory (AGT), provided the theoretical foundations for the design of *Gravitas*; they are not directly tested as comprehensive models in the present analysis, but rather serve as the design rationale against which its empirical findings should be read.

Gravitas was developed against this backdrop as a simulation-based learning game targeting common misconceptions in Newtonian mechanics through a structured cycle of diagnostic tasks, conflict-inducing animations, simulation-based exploration, and reflection (Ryan and Deci, 2000; Elliot and McGregor, 2001; Hidi and Renninger, 2006).

At the same time, learning processes and performance in game-based environments vary substantially between learners. Differences in prior knowledge as well as motivational and affective factors (such as goal orientations, interest, and persistence) shape how students engage with games and what they achieve (Ryan and Deci, 2000; Hidi and Renninger, 2006; Shute et al., 2013; Wouters et al., 2013). To address such heterogeneity, many serious games incorporate motivational design elements such as player types or mission profiles that aim to align the experience with learners' preferences and self-conceptions. However, despite their widespread use, it remains largely unclear whether such self-selected profiles correspond to meaningful differences in actual interaction behavior or learning-related outcomes.

Gravitas integrates four motivationally framed mission profiles (Adventurer, Researcher, Tactician, and Tinkerer) which learners choose at the start of the game (Richter et al., 2026). These profiles are presented as self-selected in-game roles that symbolically reflect different orientations toward exploration, experimentation, strategy, and problem solving. Critically, however, they do not alter the core mechanics or progression of the game, and they have not been validated as psychological constructs or stable learner types. Their function in the current implementation is primarily one of narrative framing; as an analytic variable, they are best understood as self-reported orientations whose behavioral correlates remain an open empirical question. This distinction matters: expectations associated with each profile should not be read as reflecting stable dispositional differences but as hypotheses about whether a lightweight motivational design element of this kind carries any predictive signal at all.

This raises a question with direct implications for the design of adaptive serious games: do self-selected profiles of this kind capture meaningful differences in how learners engage with a simulation-based environment, or do they merely provide a narrative framing without explanatory value for behavior and performance? Using log data and in-game reflection scores from $N = 272$ lower secondary students, we examine whether mission profile selection is associated with systematically different interaction patterns and whether it relates to differences in in-game reflection performance. The primary contribution is empirical: we provide evidence on whether a widely used motivational design element differentiates learner behavior and in-game conceptual performance in a simulation-based physics game. Secondary implications concern the methodological value of interaction log data for detecting such differences and the design question of what role profile-based elements can and cannot play in adaptive support systems.

RQ1: Are self-selected mission profiles associated with systematically different interaction patterns in Gravitas, as reflected in log-based behavioral indicators?

RQ2: Do self-selected mission profiles relate to differences in in-game reflection performance?

2. Theoretical Background

2.1 Simulation- and Game-based Learning in Physics

Simulation environments such as PhET (Wieman et al., 2008, 2010) provide interactive visualizations of physical systems that allow learners to manipulate parameters and observe resulting changes, supporting conceptual understanding particularly when embedded in guided instructional contexts. Recent reviews confirm that interactive simulations can significantly enhance students' conceptual understanding across physics topics when integrated into active learning environments (Banda and Nzabahimana, 2021), and that combining virtual and hands-on experimentation tends to be more effective than either approach alone (Wörner et al., 2022). Serious games extend this paradigm by embedding simulations in goal-directed scenarios with narrative framing and feedback loops. Meta-analytic evidence indicates positive cognitive and motivational effects compared to conventional instruction, particularly when instructionally well aligned and scaffolded (Wouters et al., 2013). In physics education, Shute et al. (2013) demonstrate that Newton's Playground supports learning in qualitative physics while providing rich process data for assessing problem solving and conceptual understanding.

Gravitas sits at the intersection of these strands: a simulation-based learning game for lower secondary physics integrating structured diagnostic items, conflict induction, and exploratory simulations to address robust misconceptions about gravity and forces. Each level follows a recursive cycle of prediction, conflict induction, concept-focused feedback, simulation-based exploration, and a post-question with optional repetition. The current study focuses on observed interaction patterns and in-game reflection performance within this design rather than on testing conceptual change theory comprehensively (Richter et al., 2026).

2.2 Conceptual Change, Diagnosis, and Motivational Design

Conceptual change theory emphasizes that learners' prior conceptions are coherent and resistant to change; instruction must make them visible, induce cognitive conflict, and support construction of more coherent scientific models (Chi, 2008; Posner et al., 1982; Limón, 2001; Vosniadou, 2013). *Gravitas* implements these principles through FCI-inspired diagnostic items, conflict-inducing animations that directly contradict intuitive beliefs, and immediate concept-specific non-evaluative feedback (Richter et al., 2026). For diagnosis, it draws on Knowledge Space Theory (Doignon and Falmagne, 1985), formalizing misconceptions in a structure-oriented model allowing inferences about latent knowledge states. Event-level logging captures answers, response times, simulation use, and navigation, enabling analyses of learning trajectories (Richter et al., 2026)

– consistent with evidence that in-game logs reliably indicate knowledge acquisition and predict post-game outcomes (Alonso-Fernández et al., 2020).

The motivational architecture draws on Self-Determination Theory (Ryan and Deci, 2000) and Achievement Goal Theory (Elliot and McGregor, 2001): no time pressure, leaderboards, or punitive evaluation; errors are framed as productive entry points for reflection. In this paper, these theories serve as design rationales rather than directly tested models.

2.3 Mission Profiles and the Open Question of Profiling Validity

Gravitas introduces four motivationally framed mission profiles (Adventurer, Researcher, Tactician, Tinkerer) chosen by learners at the start of the game. They do not influence core mechanics or progression but provide narrative framing and serve as an a priori grouping variable. Analyses of interaction logs in simulation-based environments have identified distinct behavioral clusters associated with different learning outcomes, suggesting meaningful patterns can be recovered from process data (Cock et al., 2025). Gameplay behavior traces have further been shown to predict posttest performance and interest, with potential for real-time adaptive scaffolding (Emerson et al., 2020). Nevertheless, rigorous empirical validation of self-selected profiles as predictors of interaction patterns and performance remains scarce in school-based science education.

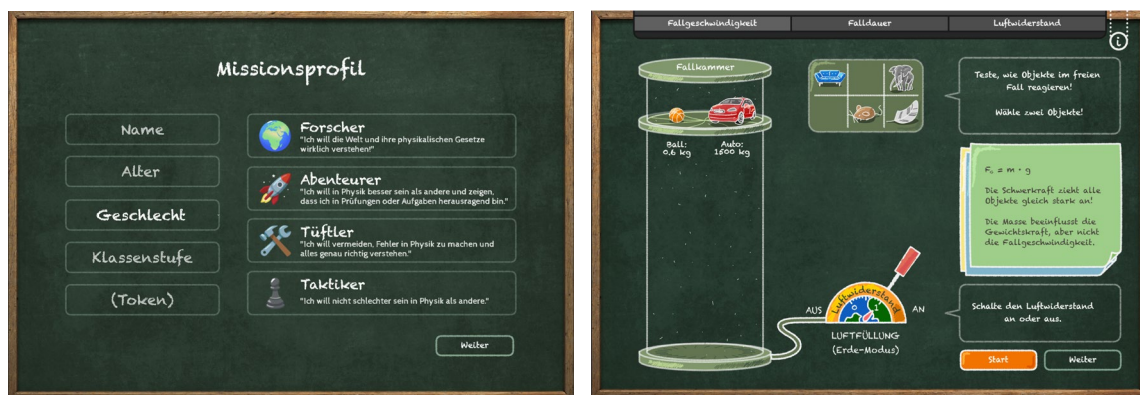


Figure 1: Screenshots from *Gravitas*. Left: mission profile selection screen, where learners choose one of four motivationally framed roles without affecting game mechanics or progression. Right: simulation screen from Level 2 (falling chamber), where learners manipulate object parameters and observe fall trajectories, generating the core interaction data analysed in this study

The mission profiles are treated as proxies for learner orientations – toward exploration, experimentation, strategic optimization, or tinkering – that should shape how students appropriate the game's affordances (Hou et al., 2021; Camacho and Nunes, 2025). Adaptive systems leveraging log-based behavioural data have produced stronger learning efficiency gains than non-adaptive versions (Debeer et al., 2021). Accordingly, we expected Adventurers and Tinkerers to show higher simulation time and reset rates, Researchers to access conceptual explanations more frequently, and Tacticians to progress efficiently with fewer resets (Lavoué et al., 2019; Halifax et al., 2020).

3. Method

3.1 Participants and Design

Participants were $N = 272$ lower-secondary students who interacted with *Gravitas* in the context of regular physics lessons. The sample had a mean age of $M = 14.22$ years ($SD = 0.97$, range = 11–17). In terms of gender, 143 students identified as male (52.6%), 111 as female (40.8%), and 18 as diverse (6.6%).

The study employed a quasi-experimental, non-interventionist design. At the beginning of the game, learners self-selected one of four mission profiles, which served as the grouping variable for all subsequent analyses. The distribution across profiles was as follows: Adventurer (34.2%), Researcher (36.4%), Tactician (19.9%), and Tinkerer (9.6%).

In line with the study's focused research questions, mission profiles were not treated as validated psychological scales but as self-selected, motivationally framed in-game categories that allowed us to compare interaction patterns and in-game reflection performance across groups.

3.2 Learning Environment: Gravitas

Gravitas is a linear, scenario-based learning game consisting of six content-focused levels, each targeting a specific misconception in mechanics derived from the FCI and related research (e.g., “heavier objects fall faster,” “gravity ceases in space,” “rockets need air to function”). Each level implements a five-phase microcycle: pre-question (diagnosis), conflict induction, feedback, simulation, and post-question with optional repetition. The simulations are deliberately implemented as simplified 2D environments to maintain transparency of causal relations and reduce extraneous cognitive load (Sweller, 1994). The game avoids high scores, leaderboards, and time pressure; summary feedback on correct answers and time spent is provided solely for self-monitoring rather than summative assessment, reflecting the SDT- and AGT-informed design rationale described above.

3.3 Procedure

Students used *Gravitas* during several consecutive class periods without strict time limits, allowing individual differences in game progress. They selected a mission profile at the start and then worked through as many levels as they could at their own pace. All interactions, responses to pre- and post-questions, simulation use, resets, access to conceptual rollups, reflection times, and navigation events, were logged with timestamps. Because not all learners completed all six levels, game progress was operationalized as the number of fully completed levels per student. To avoid confounding behavioral measures with differing amounts of progress, interaction variables were normalized per completed level.

3.4 Measures

Independent variable. Mission profile (Adventurer, Researcher, Tactician, Tinkerer) was included as a self-selected categorical factor.

Interaction measures (per completed level). To account for differences in overall task exposure, all behavioral indicators were normalized by the number of completed levels (C_i), yielding per-level metrics. This normalization approximates the average intensity of interaction per unit of task progression.

Four categories of interaction were operationalized:

Exploration intensity. Exploration behavior was captured as the mean simulation time per completed level:

$$\text{SimulationDuration}_i = \frac{\sum_{l \in C_i} \text{Duration}_{il}}{|C_i|}$$

where Duration_{il} denotes the time learner i spent in simulation in level l .

Trial-and-error behavior. Trial-based exploration was operationalized as the mean number of simulation resets across levels with available reset indicators:

$$\text{ResetFrequency}_i = \frac{\sum_{l \in L_i} \text{Resets}_{il}}{|L_i|}$$

where Resets_{il} denotes the number of resets of learner i in level l , and L_i is the subset of levels for which reset data were available. Because reset indicators were only recorded for a subset of levels, normalization was based on $|L_i|$ rather than the total number of completed levels.

Use of conceptual supports. Engagement with embedded conceptual explanations (“rollups”) was measured as the mean number of rollup visits per completed level:

$$\text{RollupVisits}_i = \frac{\sum_{l \in C_i} \text{Visits}_{il}}{|C_i|}$$

where Visits_{il} denotes the number of rollup accesses by learner i in level l .

Reflection engagement. Reflective processing was operationalized as the mean time spent in reflection phases per completed level:

$$\text{ReflectionDuration}_i = \frac{\sum_{l \in C_i} \text{ReflectionTime}_{il}}{|C_i|}$$

where $\text{ReflectionTime}_{il}$ denotes the time learner i spent in reflection in level l .

For variables with incomplete level coverage (e.g., reset indicators), normalization was based on the number of levels with available data. Additional indicators (e.g., number of attempts and first-try correctness) were included in the broader analytical framework but are not reported in detail here.

In-game conceptual performance. Reflection performance was operationalized as the mean score on post-simulation reflection tasks per completed level. The number of completed levels was treated as a progress indicator but not included as an independent predictor due to its conceptual entanglement with opportunities to earn reflection points.

3.5 Data Analysis

Group differences between mission profiles were analyzed using non-parametric Kruskal-Wallis tests, appropriate given the skewed distributions of several log variables, with ϵ^2 calculated for each dependent variable. Associations between gameplay behavior and in-game conceptual performance were examined using Spearman rank correlations (pairwise complete observations). Given the exploratory nature of the study, no correction for multiple testing was applied; results are interpreted with emphasis on effect sizes and overall patterns.

4. Results

4.1 Profile Distribution and Game Progress

The sample was unequally but adequately distributed across mission profiles: Adventurer (n = 93), Researcher (n = 99), Tactician (n = 54), and Tinkerer (n = 26). On average, learners completed close to five levels (M = 5.00, SD = 1.48). Table 1 provides an overview of descriptive statistics and test results for all behavioural indicators across mission profiles. The number of completed levels did not differ significantly between profiles (Table 1). Descriptively, Tinkerers showed the highest progress, followed by Adventurers, Tacticians, and Researchers, but these differences were negligible.

Table 1: Descriptive statistics and profile comparison for behavioural indicators and reflection performance across mission profiles

	Adventurer (n = 93)	Researcher (n = 99)	Tactician (n = 54)	Tinkerer (n = 26)	$\chi^2(3)$	p	ϵ^2
Levels completed	M = 5.14, SD = 1.36	M = 4.74, SD = 1.71	M = 5.00, SD = 1.49	M = 5.46, SD = 0.51	2.16	.540	.000
Simulation duration (s/level)	M = 61.6, SD = 101.9	M = 85.1, SD = 184.9	M = 68.8, SD = 75.8	M = 72.4, SD = 92.5	1.76	.624	.000
Reset frequency	M = 3.05, SD = 2.15	M = 4.16, SD = 7.17	M = 3.78, SD = 2.19	M = 3.67, SD = 1.87	7.59	.055	.019
Rollup visits (per level)	M = 0.20, SD = 0.57	M = 0.24, SD = 0.93	M = 0.23, SD = 0.49	M = 0.21, SD = 0.29	1.28	.734	.000
Reflection duration (s/level)	M = 27.7, SD = 20.9	M = 38.5, SD = 42.9	M = 31.5, SD = 20.7	M = 37.0, SD = 40.3	5.69	.128	.010
Reflection performance (pts/level)	M = 57.98, SD = 7.83	M = 58.03, SD = 10.03	M = 58.14, SD = 7.59	M = 56.96, SD = 7.24	0.68	.879	.000

Note. Kruskal-Wallis tests. ϵ^2 = epsilon-squared effect size. Bold: marginal effect for reset frequency (p = .055).

4.2 Interaction Patterns in the Simulation

Exploration intensity (simulation duration per completed level). No significant profile differences emerged in exploration intensity (Table 1). Descriptively, Researchers spent the most time in simulations, whereas Adventurers showed the shortest durations; Tacticians and Tinkerers lay in between. The effect size indicates these differences were trivial.

Trial-and-error behavior (reset frequency). Reset frequency showed a marginally significant profile difference, $\chi^2(3) = 7.59$, $p = .055$, $\epsilon^2 = .019$ – the only indicator approaching significance. Researchers (M = 4.16, SD = 7.17) and Tacticians (M = 3.78, SD = 2.19) reset simulations more frequently than Adventurers (M = 3.05, SD = 2.15), with Tinkerers falling in between (M = 3.67, SD = 1.87). The comparatively large standard deviation among Researchers suggests substantial within-profile heterogeneity, with a subset of learners engaging in

particularly intensive trial-and-error behavior. Given the marginal significance level and small effect size, this pattern should be interpreted cautiously.

Use of conceptual support (rollup visits per completed level). No significant profile differences emerged in rollup use (Table 1). Average visits per level were low across all groups, with Researchers showing the highest and Adventurers the lowest values. The effect size indicates no meaningful differentiation.

Reflection duration per completed level. Reflection duration did not differ significantly across profiles (Table 1). Descriptively, Researchers and Tinkerers spent somewhat more time in reflection phases than Adventurers and Tacticians, but the effect remained very small.

Taken together, the mission profiles did not produce robustly distinct interaction patterns. Only reset behavior showed a weak tendency toward differentiation, and even this effect was small.

4.3 Learning Outcomes and Relations to Behavior

Profile differences in reflection performance. Reflection performance did not differ significantly between mission profiles (Table 1). Average reflection scores were highly similar across groups, ranging from $M = 56.96$ to $M = 58.14$ points, with Tacticians scoring highest and Tinkerers lowest. Thus, self-selected mission profiles were not associated with meaningful differences in in-game conceptual performance.

Correlations between interaction measures and reflection performance. Spearman correlations revealed no significant associations between reflection performance and simulation duration ($p = -.03, p = .628$), rollup use ($p = -.02, p = .713$), or reflection duration ($p = .01, p = .811$). In contrast, reset frequency showed a small but statistically significant negative association with reflection performance ($p = -.18, p = .006$). Learners who reset simulations more frequently therefore tended to achieve slightly lower reflection scores. While statistically reliable, this association was small in magnitude.

Overall, self-selected mission profiles were not linked to differentiated learning outcomes, and most interaction measures were unrelated to reflection performance. The only consistent signal was a weak negative relation between reset frequency and conceptual performance within the game.

5. Discussion

5.1 Summary of Main Findings

This study examined whether self-selected, motivationally framed mission profiles in the simulation-based physics game *Gravitas* are associated with distinct interaction patterns and differences in in-game reflection performance. Across a sample of 272 secondary students, no meaningful profile differences emerged in game progress, exploration intensity, use of conceptual support, reflection duration, or reflection performance. Corresponding effect sizes were negligible, indicating that mission profiles did not differentiate learners' interaction behavior or in-game performance outcomes.

Only reset behavior showed a marginal profile effect, with Researchers and Tacticians exhibiting somewhat higher reset frequencies than Adventurers. The effect was small and accompanied by high within-profile variability. Independent of profile, reset frequency showed a small negative correlation with reflection performance ($p = -.18$), suggesting that more frequent resetting was weakly associated with lower reflection scores. Given the correlational nature of the data, no conclusions can be drawn about underlying cognitive processes.

Simulation duration, reflection duration, and use of conceptual supports were not associated with reflection performance. Thus, neither increased time on task nor engagement with available support translated into higher in-game conceptual scores.

Taken together, these findings indicate that self-selected mission profiles in *Gravitas* have limited explanatory value for learners' interaction patterns and in-game performance. These conclusions are restricted to log data and reflection performance and should not be generalized to broader learning outcomes without additional evidence.

5.2 Relation to Prior Work

The absence of robust behavioral and performance differences between mission profiles challenges the assumption that self-selected player types align with distinct interaction styles and in-game learning outcomes. More broadly, the findings are consistent with research suggesting that learning behavior in simulation-based

environments is dynamic and context-dependent. Individual differences tend to emerge in interaction trajectories, such as error patterns, exploration strategies, and help-seeking behavior, rather than in static self-descriptions (Shute et al., 2013; Wouters et al., 2013; Bodily and Verbert, 2017).

Accordingly, the present findings do not support the use of self-selected mission profiles as stand-alone indicators for adaptive support decisions in *Gravitas*. This interpretation aligns with prior work showing that behavior-based or hybrid profiling approaches yield more reliable predictions of engagement and learning outcomes than static player-type labels (Lavoué et al., 2019; Hallifax et al., 2020; Rodríguez et al., 2022). Evidence from game-based learning research further indicates that in-game process indicators, including performance trajectories and affective states, are more closely associated with learning processes than stable categorical profiles (Shute et al., 2021; Kim and Shute, 2015).

The absence of profile effects should not be interpreted as an absence of individual differences but rather as an indication that such differences manifest through the interaction of learner states, affective dynamics, and task demands.

From a Self-Determination Theory perspective, the autonomy-supportive design of *Gravitas* may have attenuated behavioral differences between learners. From an Achievement Goal Theory perspective, mission profiles may not validly capture the motivational or strategic dispositions they are intended to represent.

5.3 Design Implications

From a design perspective, the findings indicate that self-selected mission profiles provide little basis for adaptive decision-making in their current form and should not be used as stand-alone drivers of adaptivity. Instead, adaptive systems in serious games should rely more strongly on behavior-based indicators that capture learners' interaction dynamics in situ.

More promising approaches include deriving profiles from observed interaction patterns (e.g., via clustering), combining self-reported orientations with behavioral indicators, and using real-time analytics to detect potentially maladaptive patterns during gameplay.

In *Gravitas*, the negative association between reset frequency and reflection performance illustrates such a behaviorally grounded indicator. Although the effect size is small, frequent resetting may signal unstructured trial-and-error strategies without sufficient conceptual consolidation. For example, high reset rates could trigger prompts encouraging prediction and explanation before further interaction.

Beyond in-game adaptivity, teacher-facing dashboards based on log data could make such behavioral patterns visible at the classroom level and support targeted instructional guidance. In this context, mission profiles may still serve as a descriptive layer, but instructional decisions should primarily be informed by observed interaction patterns and performance indicators.

5.4 Limitations and Future Research

Several limitations should be considered. First, the construct validity of the mission profiles remains uncertain: profiles were based on a single self-selection and were not validated against independent measures of motivational orientation or strategic tendencies. Profile choice may therefore reflect situational preference or self-presentation rather than stable learning-relevant dispositions. Future research should examine how mission profiles relate to established constructs such as goal orientations, interest development, or self-regulated learning.

Second, the outcome measure was restricted to in-game reflection performance. While closely aligned with the instructional design of *Gravitas*, this measure does not capture broader conceptual understanding, transfer, or long-term learning effects (Richter et al., 2026). External assessments such as FCI-based measures or misconception interviews would allow for triangulation.

Third, the analyses did not control for confounding variables including prior physics knowledge, gameplay exposure, class-level or teacher effects, familiarity with digital games, or differences in reading and digital interaction skills – factors that may partly account for the absence of significant profile differences. Future studies should employ multilevel modelling including such covariates.

Fourth, generalization beyond lower-secondary physics in the Swiss context should be made cautiously (Richter et al., 2026). Future research should focus on modelling learning behaviour dynamically, with behavior-based or hybrid profiling approaches, real-time adaptive systems, and longitudinal designs.

6. Conclusion

This study examined whether self-selected mission profiles in the simulation-based physics game *Gravitas* are associated with distinct interaction patterns (RQ1) and differences in in-game reflection performance (RQ2). The findings consistently indicate that they are not: across all behavioural indicators and performance measures, profile membership showed negligible explanatory value. The only exception was a marginal and small effect for reset frequency. These results contribute empirical evidence that lightweight motivational design elements of this kind do not reliably differentiate learner behaviour or in-game conceptual performance in simulation-based physics games, and should not serve as stand-alone drivers of adaptive support.

Ethics Declaration: This study was conducted in accordance with applicable institutional and national ethical guidelines for educational research. According to the relevant institutional regulations, formal ethical approval was not required for non-invasive studies conducted in regular school settings. Data collection was coordinated through school administrations, which informed and authorized participating teachers. Participation was voluntary, and all data were anonymised before analysis. Individual log records were linked to numerical identifiers only and could not be traced back to specific persons.

AI Declaration: AI tools (e.g., ChatGPT, OpenAI) were used to support academic editing, linguistic refinement, and formatting during manuscript preparation. No AI tools were used to generate or analyse research data. All content was critically reviewed and verified by the authors to ensure accuracy and integrity

References

- Alonso-Fernández, C. et al (2020) "Predicting students' knowledge after playing a serious game based on learning analytics data: A case study", *Journal of Computer Assisted Learning*, Vol 36, No. 3, pp 350–358.
- Banda, H.J. and Nzabimana, J. (2021) "Effect of integrating physics education technology simulations on students' conceptual understanding in physics: A review of literature", *Physical Review Physics Education Research*, Vol 17, No. 2, p 023108.
- Bodily, R.G. and Verbert, K. (2017) "Review of research on student-facing learning analytics dashboards and educational recommender systems", *IEEE Transactions on Learning Technologies*, Vol 10, pp 405–418.
- Camacho, É. and Nunes, P. (2025) "Building player profiles for strategic analysis in higher education", *Journal of Entrepreneurial Researchers*, Vol 3, No. 1.
- Chi, M.T.H. (2008) "Three types of conceptual change", in Vosniadou, S. (Ed.), *Handbook of Research on Conceptual Change*, Erlbaum, pp 61–82.
- Cock, J.M. et al (2025) "Different topics, similar patterns: The consistency of learners and learning across interactive STEM simulations", *International Journal of Artificial Intelligence in Education*.
- Debeer, D. et al (2021) "The effect of adaptivity in digital learning technologies: Modelling learning efficiency using data from an educational game", *British Journal of Educational Technology*, Vol 52, No. 5, pp 1881–1897.
- Doignon, J.-P. and Falmagne, J.-C. (1985) "Spaces for the assessment of knowledge", *International Journal of Man-Machine Studies*, Vol 23, No. 2, pp 175–196.
- Elliot, A.J. and McGregor, H.A. (2001) "A 2×2 achievement goal framework", *Journal of Personality and Social Psychology*, Vol 80, No. 3, pp 501–519.
- Emerson, A. et al (2020) "Multimodal learning analytics for game-based learning", *British Journal of Educational Technology*, Vol 51, No. 5, pp 1505–1526.
- Hallifax, S. et al (2020) "To tailor or not to tailor gamification?", *Lecture Notes in Computer Science*, Vol 12163, pp 216–227.
- Halloun, I.A. and Hestenes, D. (1985) "The initial knowledge state of college physics students", *American Journal of Physics*, Vol 53, No. 11, pp 1043–1055.
- Hidi, S. and Renninger, K.A. (2006) "The four-phase model of interest development", *Educational Psychologist*, Vol 41, No. 2, pp 111–127.
- Hou, X. et al (2021) "Assessing the effects of open models of learning and enjoyment in a digital learning game", *International Journal of Artificial Intelligence in Education*.
- Kim, Y.J. and Shute, V.J. (2015) "Opportunities and challenges in assessing and supporting creativity in video games", in Green, G.P. and Kaufman, J.C. (Eds.), *Video Games and Creativity*, Elsevier Academic Press, pp 99–117.
- Lavoué, E. et al (2019) "Adaptive gamification for learning environments", *IEEE Transactions on Learning Technologies*, Vol 12, No. 1, pp 16–28.
- Limón, M. (2001) "On the cognitive conflict as an instructional strategy", *Learning and Instruction*, Vol 11, No. 4–5, pp 357–380.
- McCloskey, M. et al (1983) "Intuitive physics: The straight-down belief", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol 9, No. 4.
- Posner, G.J. et al (1982) "Accommodation of a scientific conception", *Science Education*, Vol 66, pp 211–227.

- Richter, K., Tschirky, D. and Kickmeier-Rust, M. (2026) "Gravitas: A simulation-based learning game for diagnosing misconceptions in mechanics", in Bakkes, S. et al (Eds.), Games and Learning Alliance. GALA 2025, Lecture Notes in Computer Science, Vol 16307, Springer, Cham.
- Rodríguez, I., Puig, A. and Rodríguez, À. (2022) "Towards adaptive gamification: A method using dynamic player profile and a case study", Applied Sciences, Vol 12, No. 1, p 486.
- Ryan, R.M. and Deci, E.L. (2000) "Intrinsic and extrinsic motivations", Contemporary Educational Psychology, Vol 25, No. 1, pp 54–67.
- Shute, V.J., Ventura, M. and Kim, Y.J. (2013) "Assessment and learning of qualitative physics", Journal of Educational Psychology, Vol 106, No. 6, pp 423–430.
- Shute, V. et al (2021) "Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games", Journal of Computer Assisted Learning, Vol 37, No. 1, pp 127–141.
- Sweller, J. (1994) "Cognitive load theory, learning difficulty, and instructional design", Learning and Instruction, Vol 4, No. 4, pp 295–312.
- Vosniadou, S. (2013) International Handbook of Research on Conceptual Change, Taylor and Francis.
- Wieman, C.E., Adams, W.K. and Perkins, K.K. (2008) "PhET: Simulations that enhance learning", Science, Vol 322, No. 5902, pp 682–683.
- Wieman, C.E. et al (2010) "Teaching physics using PhET simulations", The Physics Teacher, Vol 48, No. 4, pp 225–227.
- Wouters, P. et al (2013) "A meta-analysis of serious games", Journal of Educational Psychology, Vol 105, No. 2, pp 249–265.
- Wörner, S., Kuhn, J. and Scheiter, K. (2022) "The best of two worlds: A systematic review on combining real and virtual experiments in science education", Review of Educational Research, Vol 92, No. 6, pp 911–952.