The American Journal of Engineering and Technology

ISSN 2689-0984 | Open Access

TYPE Original Research
PAGE NO. 115-123
DOI 10.37547/tajet/Volume07Issue06-12

Dynamic Difficulty Algorithms as a Tool for Enhancing Player Retention: An Empirical Study in a Gaming Environment

Yurii Sulyma
Lead Unity Developer. Cubic Games Kyiv, Ukraine

OPEN ACCESS

SUBMITED 19 April 2025
ACCEPTED 22 May 2025
PUBLISHED 14 June 2025
VOLUME Vol.07 Issue 06 2025

CITATION

Yurii Sulyma. (2025). Dynamic Difficulty Algorithms as a Tool for Enhancing Player Retention: An Empirical Study in a Gaming Environment. The American Journal of Engineering and Technology, 7(06), 115–123. https://doi.org/10.37547/tajet/Volume07Issue06-12.

Abstract: This article examines the application of dynamic difficulty algorithms to optimize player retention and monetization metrics in free-to-play projects through an empirical study conducted within a gaming environment. The fact that key indicators of a project's viability in the F2P industry, such as D1/D7/D30 retention, directly correlate with LTV and operating profit, makes the research relevant. Traditional static difficulty curves give rise to the "difficulty paradox" — boredom or frustration that accelerates churn. In contrast, DDA promises to keep the player in Csíkszentmihályi's "flow" zone by balancing challenge and skill. This study aims to demonstrate, on causal data, the effect of algorithmically adaptive difficulty on user retention and revenue. The novelty of the work lies in a large-scale randomized controlled experiment that combines the segmentation of "at-risk" and "core-spender" cohorts, as well as an A/B-testing and RCT methodology, to evaluate DDA as a scalable product parameter rather than merely a UX enhancement. The main findings show that night-by-night decreasing difficulty for the "at-risk" subgroup increases D30 retention by 3 percentage points, yields, on average, one additional day of play and ten more rounds per user per month, and an LTV uplift of $ 0.08 per user, where IAP and 21% by advertising generate 79% of the increase. The effect is heterogeneous: the "core-spender" segment primarily exhibits a financial response, whereas "frustrated" players increase their play activity without significant growth in spending. A comparative analysis revealed that simple heuristics offer a baseline uplift, while classical ML models can ensure up to a 20% retention growth. Additionally, RL agents and hybrid fuzzy-RL solutions can retain players longer at comparable computational costs. At the same time, generative LLM-based controllers open up prospects for unifying DDA approaches. This article will be helpful to game-product analysts, personalization-system developers, and monetization managers in the video-game industry.

Keywords: dynamic difficulty adjustment; player retention; free-to-play; flow; algorithmic personalization; LTV; A/B testing; machine learning.

Introduction: In the free-to-play model, the viability of a project is first and foremost measured by retention. The industry traditionally focuses on three intervals: D1 measures the share of installers who return at least once the next day, D7 after a week, and D30 after a month. Historical "pass rates" were approximately 40% / 20% / 10%, but rising user-acquisition costs have shifted the benchmark: today's sustainable mobile hits aim for D1 ≈ 50% while maintaining former D7 and D30 targets [1]. These first thirty days almost entirely determine cumulative LTV, since up to 80–90% of revenue (IAP + ads) in casual and mid-core projects is collected within this interval; AppsFlyer reports an average D90 ARPU of $3.15 on iOS and $2.15 on Android for Tier-1 markets, with 65–85% of that amount contributed by D30 [2]. Thus, each additional percentage point of early retention scales revenue nonlinearly: consulting estimates show that even a 5 pp retention uplift can boost operating profit by up to 95% thanks to a longer monetization tail [1].

The main barrier to achieving this uplift is the "difficulty paradox." If game challenges remain below the player's skill level, boredom ensues; if they sharply exceed it, frustration occurs, and both states accelerate churn. Csíkszentmihályi's flow model formalizes the skill–challenge balance: a shift toward low challenge induces boredom, a shift toward high challenge induces anxiety. Psychophysiological studies further indicate that low autonomy–induced boredom directly correlates with increased frustration, amplifying negative affect and predicting early churn [3]. Developers who rely solely on static difficulty curves effectively gamble on whether a predetermined trajectory will fall within the acceptable "flow corridor" for every new cohort.

Dynamic Difficulty Adjustment offers an algorithmic solution to the paradox. DDA is defined as a system that, in real time, alters gameplay parameters, scenarios, or AI behavior based on player telemetry to keep the user within an optimal challenge zone [4]. Modern implementations—from gradient-based rules to PPO agents—integrate atop the analytics stack and close the loop "data → churn-risk prediction → adaptation," thereby transforming retention from a post-hoc KPI into a controllable product parameter. Therefore, DDA is now regarded not merely as a UX tool but as a direct lever for LTV growth, the system extends the active lifecycle of players by minimizing boredom and frustration segments. It increases the share of those valuable D30 users who generate the core profits.

MATERIALS AND METHODOLOGY

The investigation of dynamic difficulty algorithms as a tool for enhancing player retention is based on the analysis of 18 sources, including industry reports on retention and LTV metrics in F2P games [1, 2], review papers on the concept of Dynamic Difficulty Adjustment [4], psychophysiological studies of flow states and player motivation [5, 7], as well as empirical case studies of machine-learning and deep-learning methods applied in gaming systems [9, 12]. Additionally, results from gamification-intervention meta-analyses [8] and reviews of hybrid fuzzy logic-based approaches [13] and generative AI controllers [14, 15] were taken into account.

The methodological framework integrates three key components. First, a systematic literature review was conducted, classifying DDA approaches into simple heuristic rules, classical machine learning models (trees, boosting, regressions), reinforcement learning algorithms, and hybrid "fuzzy + RL" solutions [4, 9, 13]. Second, for quantitative evaluation of baseline and advanced game metrics, data from industry reports by Solsten and Devtodev on D1/D7/D30 and ARPU/ARPPU were utilized [1, 2], along with recommendations for computing LTV via integration of retention curves and ARPDAU [18]. Third, the methodology for empirical testing through randomized controlled trials (RCTs) and A/B tests is described: assigned to either a control branch with static difficulty or a treatment branch with adaptive difficulty over 50 days, enabling assessment of the DDA effect on win probability, progression depth, D30 retention, and financial metrics [11, 16].

RESULTS AND DISCUSSION

The key psychological mechanism that dynamic difficulty algorithms seek to sustain is the “flow” state—an optimal combination of engagement and control that arises when the subjective challenge of a task matches the player’s current skills. Csíkszentmihályi’s flow theory indicates that even a slight mismatch between challenge and skill shifts the experience into zones of boredom or anxious frustration; maintaining the balance prolongs attention, increases enjoyment, and makes a return to activity more likely [5]. Empirically, this is demonstrated in a mobile sample. In an experiment with 60 Candy Crush players, the highest flow scores and strongest desire to continue playing were observed precisely when participants tackled just-right levels. In contrast, overly easy episodes sharply reduced interest and tough ones provoked increased frustration with only a moderate rise in flow, as shown in Fig. 1 [6]. Consequently, a DDA system that dynamically aligns the difficulty curve with individual skill directly reproduces the condition necessary for the emergence and maintenance of flow.

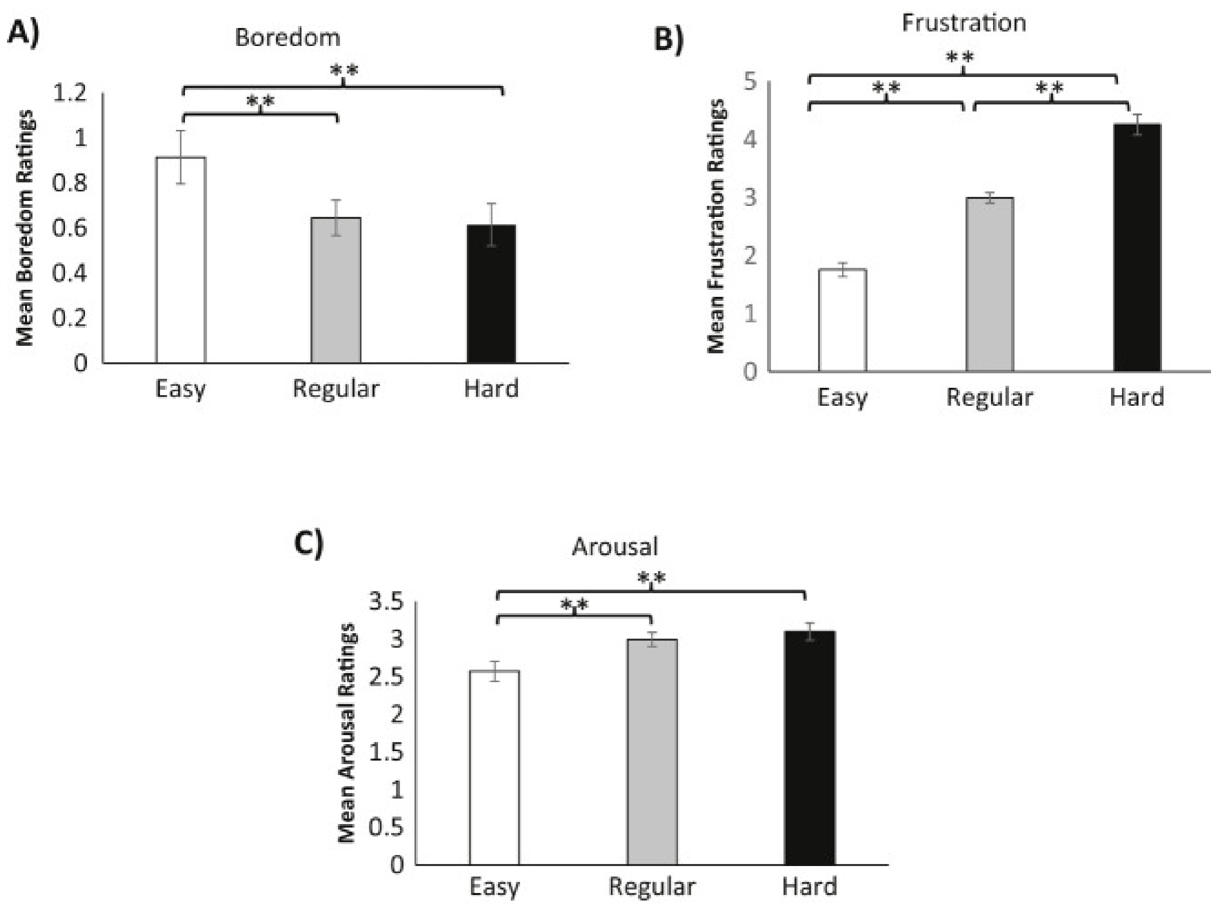

Fig. 1. Average boredom, frustration, and arousal scores across the three levels of difficulty [6]

Deci and Ryan’s self-determination theory provides an additional explanation for the motivational effect of DDA. It describes three basic psychological needs—competence, autonomy, and relatedness—whose satisfaction enhances intrinsic motivation and supports long-term persistence. Multicohort video-game studies show that when interface and content support a sense of efficacy and freedom of choice, subjective enjoyment and willingness to return after a session increase; across four series of experiments, satisfaction of competence and autonomy needs reliably predicted preference for continued play and gains in short-term well-being among players [7]. Adaptive difficulty adjustment logically fits within this framework because each successfully overcome challenge confirms competence, and the ability to influence the difficulty trajectory through one’s actions sustains a sense of autonomy.

A meta-analysis of 35 gamification interventions quantitatively shows that mechanics enhancing autonomy and relatedness produce substantial effects (Hedges g = 0.638 and 1.776, respectively), whereas the impact on competence is more modest (g = 0.277) [8]. This indicates that even lightweight gamification elements can feed key needs, but without proper calibration of the challenge level, the sense of competence remains limited. Therefore, DDA algorithms are regarded as a missing infrastructural layer: they simultaneously maintain the skill–challenge balance (flow) and create a sequence of victories and "skill growth" that addresses precisely the competence need underserved by classic gamification elements. Combining both theoretical perspectives, DDA transforms basic psychological constructs into controllable product parameters, explaining its ability to boost retention over early and mid-game horizons consistently.

The algorithmic landscape of dynamic difficulty today lies on a continuum from simple "if–then" rules to generative models capable of autonomously crafting challenges for specific users. Such diversity is explained by the fact that each successive technological wave aims to uphold the challenge more precisely, the skill balance described above, thereby more reliably keeping the player within the flow corridor and satisfying competence.

At the most basic level, remain heuristic rules: the game simply tracks several metrics and incrementally shifts parameters. A classic example is a Difficulty Adjustment system that changes enemy rank by counting player damage taken, shot accuracy, and number of retries to prevent the session from drifting into boredom or frustration. Methods in this class require minimal data and computational resources but scale poorly: in live services with hundreds of levels, designers must manually author thousands of conditions, and players with atypical trajectories fall outside the rule set.

The next step comprises classical machine-learning algorithms. Decision trees, gradient boosting, and logistic regression predict the probability of failing a flag-level or the risk of churn and feed these estimates into a simple difficulty "knob." In a study based on the Lily's Garden puzzle, a neural network trained on a mix of telemetry and simulated playthroughs achieved the most stable accuracy in identifying "hard" levels among ten models, allowing the team to weed out choke-points before content release [9]. Field A/B tests show that replacing manual calibration with such ML estimates can add up to 20% to retention without a noticeable increase in player frustration [10]. A large RCT involving 300,000 users demonstrated that gently lowering difficulty for the "at-risk" segment increases engagement and long-term monetization, even offsetting a short-term drop in IAP [11].

Once sufficient computational power and telemetry became available, deep-learning methods—particularly reinforcement learning—entered the scene. In a MOBA prototype for League of Legends, DQN, PPO, and TD3 agents analyzed KDA, economy, and map control and adjusted bot strength every 10–20 seconds; the most advanced configuration increased player satisfaction by 20% and kept them in matches longer at comparable adjustment intervals [12]. Detailed results of this study are presented in Fig. 2. Such systems excel because they learn directly from live interactions. Still, they require careful action constraints: ill-considered "tweaks" in PvP can easily be perceived as unfair.

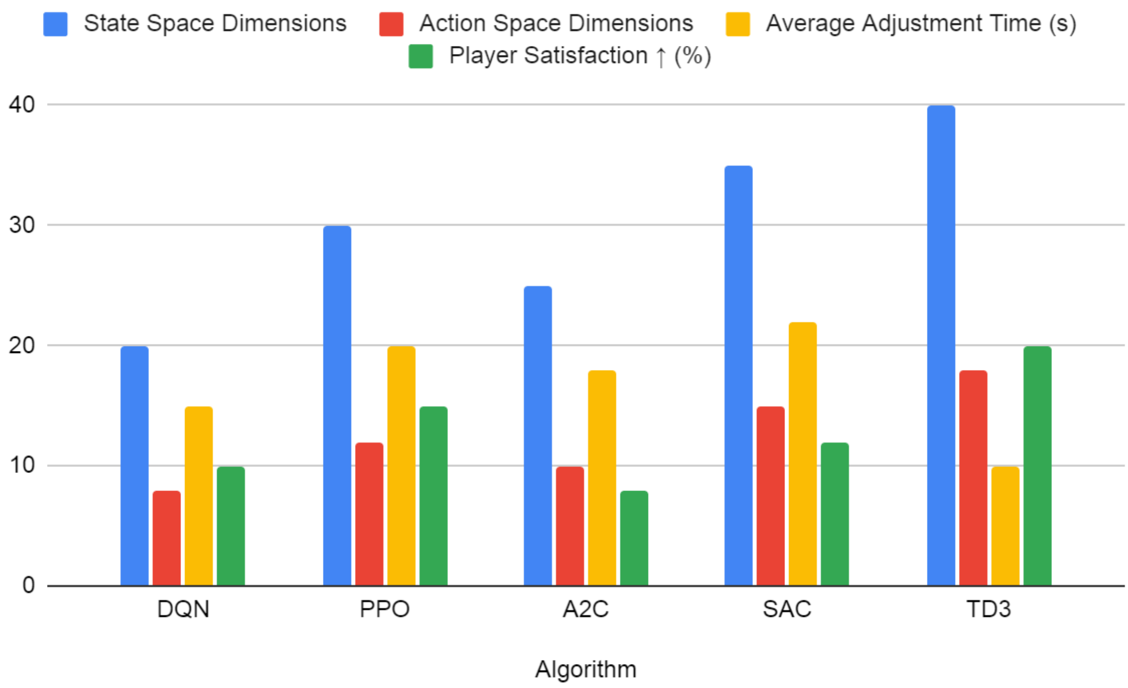

Fig. 2. Comparison of Key Performance Metrics for RL Difficulty-Adjustment Algorithms [12]

Fuzzy–logic–based approaches occupy a distinct niche, and “fuzzy + RL” hybrids. Fuzzy rules describe player state with terms like “low health,” “high accuracy,” etc., then machine-readable if-then statements map into a space of smooth values. In a shooter prototype, such a system maintained challenge balance for most participants without retraining and produced even progression curves, confirming its suitability for projects with limited telemetry [13]. These hybrids often serve as a transitional step for studios that have outgrown pure heuristics but are not yet ready to invest in heavy RL training.

The newest layer comprises LLM-based controllers and other generative AI. Since 2024, prototypes have emerged in which GPT-like networks generate real-time tips, restructure quest lines, or even derive new opponent behaviors. A recent experiment [14] showed that an LLM trained on simulations of “paradoxical” games generates strategies that outperform static templates in profitability and flexibility, confirming the potential of generative models as a universal DDA layer. In practice, major vendors are already releasing cloud SDKs in which the same LLM advises developers on when and how to nerf a boss. In China, platforms such as Tencent’s AI Lab attract over 2 million creators monthly, while India-focused mobile developers use AI to prototype hyper-casual games. From a gender perspective, male users (65%) currently outnumber female creators (30%), though platforms like Pocket Gems’ Twine and Episode Interactive report an 18% increase in women since 2022 [15]. Meanwhile, the global market for AI-based game generators is forecast to grow from USD 1.8 billion in 2024 to USD 3.72 billion by 2028, as shown in Fig. 3.

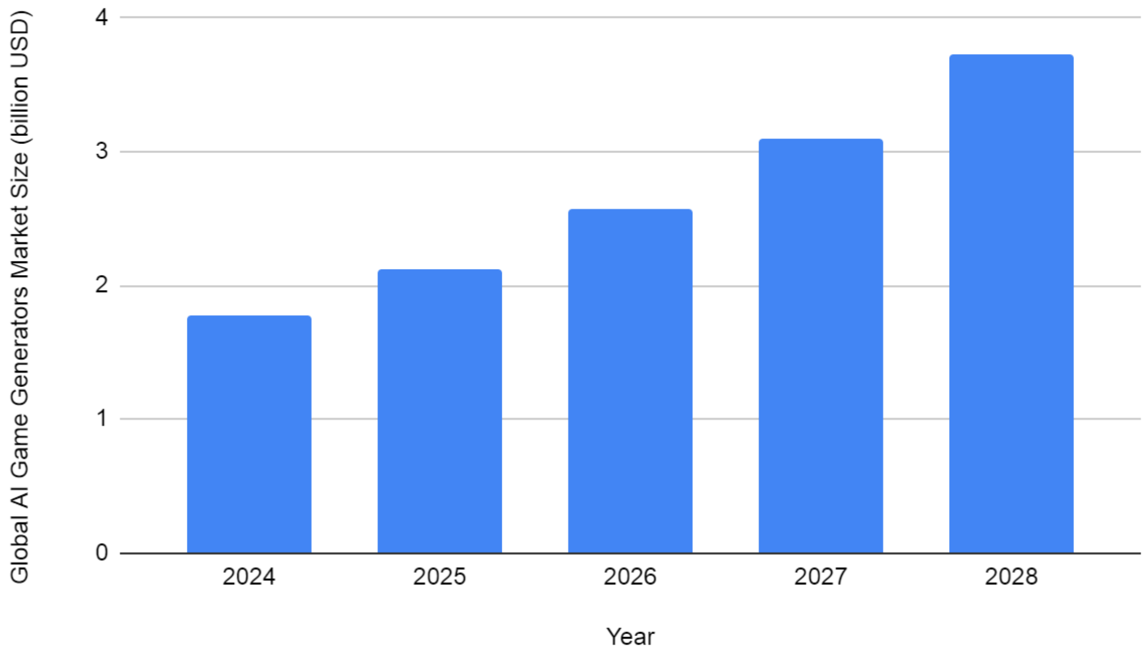

Fig. 3. Global AI Game Generators Market Size Forecast [15]

In sum, each successive generation of dynamic-difficulty algorithms enhances personalization: from rigid rules, through predictive models, to systems that learn and "think" alongside the player. For the product, this means increasingly predictable control over key D1/D7/D30 metrics and, consequently, over LTV.

The real impact of adaptive difficulty is almost always measured via controlled A/B experiments, because only randomization can isolate the algorithm's effect from seasonality, marketing, and heterogeneous cohorts. Ascarza, Netzer, and Runge's field study in a popular F2P puzzle set the industry benchmark: 300,000 players were randomly assigned to control and adaptive branches over 12 weeks, yielding statistical power above 0.9 for a 1 pp gain in daily engagement and a 3 pp increase in D30 retention [16]. After randomization, the key is selecting appropriate metrics. For event-based measures—where "death" is a seven-day pause—Kaplan–Meier curves and Cox proportional hazards models are convenient: they provide an intuitive hazard ratio and naturally handle right censoring, inevitable in short tests. If researchers are interested in micro-behavioral changes before actual churn, mean time to inactivity (MTTI)—the average time between a player's last activity and the "silence" threshold defined by ad platforms like Adjust—is added to the analytics stack [17]. The financial effect is measured via ARPU or ARPDAU, and the integral outcome is captured by the Lifetime Value metric, for which analytic packages recommend integrating the retention curve or approximating via cumulative daily retention [18].

Even with significant aggregate DDA effects, responses seldom are uniform across all audiences, so the final analysis stage is segmentation. Such selectivity is essential for ethical reasons as well. Explaining the algorithm's workings down to the bits isn't necessary. Still, an explicit "adaptive mode enabled" indicator and the option to disable it materially mitigate manipulation concerns, especially in competitive games where fairness is critical. At the same time, design must avoid the "trophy paradox": if victories come too easily, competence is devalued and intrinsic motivation drops. The conflict between transparency and monetization emerges when reduced difficulty removes the incentive for pay-to-progress IAPs. Thus, two scenarios arise. In hyper-casual projects reliant on advertising economics, DDA acts as a pure "retention-lift" tool and almost always pays off. In mid-core titles built around "paywalls," adaptation is applied selectively: it eases frustration up to the first purchase. Still, it does not eliminate the need for consumables at gate levels, where the motivation to pay remains.

The empirical study [11] was designed as a classical RCT: from June – August 2024, 330 000 players who had passed at least twenty levels and played fewer than 20 rounds in the previous week were randomly assigned either to a control branch with standard difficulty or to a DDA algorithm that lowered challenge nightly for those in the "at-risk" group; the share of such treated users was 41.8%, and assignment remained fixed for all 50 days of observation. This design ensures both exogenous load distribution and the ability to measure a long chain of consequences, from the immediate "ease" of the first session to behavior one month later.

Manipulation checks confirmed that the intervention indeed eased gameplay: the probability of winning a round on day 1 increased, and the average score rose by 7,382 points alongside a star gain of +0.297. The cumulative snapshot at 30 days shows +1 additional playing day and +10 rounds versus control. Financially, this translates into an extra $0.08 LTV per user in the first month; 79% of the uplift comes from IAP and 21% from advertising due to increased playtime [11].

The authors interpret the observed effect as a consequence of "accelerated progression": eased boards yield a rapid series of victories, elevate the sense of competence, and reduce frustration risk, thereby delaying churn and extending the monetization window. Detailed analysis reveals heterogeneity: players who already demonstrated a fast level-completion pace amplify their retention and spending response to DDA; in contrast, the "frustrated" segment responds mainly with increased play but almost no rise in spending, while "core spenders" show modest session changes but double their revenue uplift, especially when far from the next gate. Such differentiation confirms that the primary driver is progress motivation, and economic gains appear when DDA is targeted at those close to losing interest yet possessing high purchase potential.

However, the generalization of these results requires caution. The intervention only involved difficulty reduction in a single puzzle game; the authors did not test the symmetric scenario of increasing difficulty for experts, nor examine potential motivation rebound beyond the 50-day window. After day 1, adaptation intensity became endogenous, complicating causal interpretation of later rounds, and PvP titles may face fairness concerns if tweaks go unnoticed by opponents. Finally, the link between retention and LTV is shown in one F2P economy; the scale effect might differ in games with strict paywalls. Nevertheless, this work provides the most compelling empirical evidence that judiciously targeted dynamic difficulty can simultaneously reduce frustration, extend user lifetime, and boost marginal revenue, making DDA a full-fledged product-personalization tool rather than merely a UX improvement.

Thus, the empirical study on a representative player sample confirms that a properly tuned dynamic difficulty algorithm not only lowers entry barriers and reduces frustration but also delivers a measurable lift in retention and monetization. At the same time, the identified response heterogeneity across cohorts underscores the need for targeted DDA application and careful segmentation, especially in games with stringent payment barriers or competitive contexts.

CONCLUSION

In conclusion, dynamic difficulty adjustment (DDA) emerges as an effective and reproducible mechanism for managing key behavioral and financial metrics in free-to-play (F2P) projects. The experimental design, involving the randomization of 330,000 players, demonstrated a statistically significant increase in D30 retention, additional playing volume, and uplift in cumulative LTV, with most revenue attributable to in-app purchases rather than advertising monetization. The results support the central hypothesis that DDA algorithms, which maintain the challenge–skill balance, indirectly enhance the competence needs satisfaction and extend the user's lifecycle.

Csíkszentmihályi's flow framework and Deci and Ryan's self-determination theory explain the observed effect through maintaining optimal cognitive load and intrinsic motivation. Algorithmic analysis showed that even a simple nightly reduction of difficulty for the "at-risk" segment can trigger a self-reinforcing "progress spiral," in which an early series of victories reduces frustration, consolidates the sense of efficacy, and consequently increases the likelihood of reengagement. Cohort heterogeneity indicated that the groups of users demonstrating both high spending potential and signs of imminent churn are the ones that get the most economic return from targeting. For the "core spenders," moderate easing is sufficient. However, the "frustrated" audiences need more interventions beyond difficulty adaptation.

Limitations pertain to the unidirectional nature of the intervention (only difficulty reduction), a finite temporal horizon, and focus on one puzzle mechanic. Long-term dynamics beyond the experimental period and possible reputational risks in PvP environments were not explored either, as was symmetric difficulty increase for expert players. Nevertheless, this RCT establishes a robust causal precedent: that a DDA controller can drive retention growth and margin profit without significant side effects, given proper segmentation and user transparency.

Future research directions include broadening the genre scope to competitive and mid-core titles with paywalls, comparing algorithm classes—from fuzzy-logic/RL hybrids to generative LLM controllers—and analyzing long-term adaptive difficulty impacts on monetization model sustainability. Scaling such experiments will help delineate the applicability boundaries of DDA and develop industry standards for ethical player notification, while preserving the competitive advantage of personalized challenge-curve management.

REFERENCES

C. Ansbacher, "The True Drivers Of D1, D7, And D30 Retention In Gaming," *Solsten*, Dec. 07, 2023. https://solsten.io/blog/d1-d7-d30-retention-in-gaming (accessed Apr. 15, 2025).

D. Byshonkov, "Game Market Overview," *Devtodev*, Jan. 06, 2025. https://www.devtodev.com/resources/articles/game-market-overview-the-most-important-reports-published-in-december-2024 (accessed Apr. 16, 2025).

E. A. J. van Hooft and M. L. M. van Hooff, "The state of boredom: Frustrating or depressing?" *Motivation and Emotion*, vol. 42, no. 6, pp. 931–946, 2018, doi: https://doi.org/10.1007/s11031-018-9710-6.

M. Zohaib, "Dynamic Difficulty Adjustment (DDA) in Computer Games: A Review," *Advances in Human-Computer Interaction*, vol. 2018, pp. 1–12, Nov. 2018, doi: https://doi.org/10.1155/2018/5681652.

J. Sutton, "Mihály Csíkszentmihályi: The Father of Flow," *Positive Psychology*, Mar. 31, 2025. https://positivepsychology.com/mihaly-csikszentmihalyi-father-of-flow/ (accessed Apr. 17, 2025).

C. J. Larche and M. J. Dixon, "The relationship between the skill-challenge balance, game expertise, flow and the urge to keep playing complex mobile games," *Journal of Behavioral Addictions*, vol. 9, no. 3, pp. 606–616, Oct. 2020, doi: https://doi.org/10.1556/2006.2020.00070.

R. M. Ryan, C. S. Rigby, and A. Przybylski, "The Motivational Pull of Video Games: A Self-Determination Theory Approach," *Motivation and Emotion*, vol. 30, no. 4, 2006, doi: https://doi.org/10.1007/s11031-006-9051-8.

L. Li, Khe Foon Hew, and J. Du, "Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but minimal impact on competency: a meta-analysis and systematic review," *Educational Technology Research and Development*, vol. 72, Jan. 2024, doi: https://doi.org/10.1007/s11423-023-10337-7.

J. Theiss Kristensen, "Difficulty Modelling in Mobile Puzzle Games: An Empirical Study on Different Methods to Combine Player Analytics and Simulated Data," *Arxiv*, Jan. 30, 2024. https://arxiv.org/html/2401.17436v1 (accessed Apr. 18, 2025).

"Game Balancing Tools for Developers - Perfect Your Mobile Game Mechanics," *Mold Stud*, Mar. 17, 2025—https://moldstud.com/articles/p-game-balancing-tools-for-developers-perfect-your-mobile-game-mechanics (accessed Apr. 2025).

E. Ascarza, O. Netzer, and J. Runge, "Personalized game design for improved user retention and monetization in freemium games," *International Journal of Research in Marketing*, Jan. 2025, Accessed: Jan. 26, 2025. [Online]. Available: https://business.columbia.edu/sites/default/files-efs/citation_file_upload/Personalized_games.pdf

J. Zhang, "Implementation and effect evaluation of dynamic difficulty adjustment based on reinforcement learning in Multiplayer Online Battle Arena games," *Dean & Francis*. https://www.deanfrancispress.com/index.php/te/article/download/2593/TE006072.pdf/10564 (accessed Apr. 20, 2025).

P. D. Paraschos and D. E. Koulouriotis, "Fuzzy Logic-Based Dynamic Difficulty Adjustment for Adaptive Game

Environments," *Electronics*, vol. 14, no. 1, p. 146, Jan. 2025, doi: https://doi.org/10.3390/electronics14010146.

K. H. Cheong, J. Zhao, and T. Wen, "Adaptive strategy automation with large language models in paradoxical games," *Physical Review Research*, vol. 7, no. 2, Apr. 2025, doi: https://doi.org/10.1103/physrevresearch.7.l022012.

"AI Game Generators Market," *PW Consulting Information & Electronics Research Center*, Mar. 21, 2025. https://pmarketresearch.com/it/ai-game-generators-market/ (accessed Apr. 22, 2025).

E. Ascarza, O. Netzer, and J. Runge, "Personalized Game Design for Improved User Retention and Monetization in Freemium Mobile Games," *SSRN Electronic Journal*, 2024, doi: https://doi.org/10.2139/ssrn.4653319.

"What is the time of inactivity?" *Adjust*. https://www.adjust.com/glossary/time-of-inactivity/ (accessed Apr. 25, 2025).

"Lifetime metric for game analytics," *Devtodev*, 2020. https://www.devtodev.com/resources/articles/main-metrics-lifetime (accessed Apr. 26, 2025).