A Review of Player Engagement Estimation in Video Games: Challenges and Opportunities

AMMAR RASHED and SHERVIN SHIRMOHAMMADI, School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Ontario, Canada
IHAB AMER, Advanced Micro Devices Inc, Markham, Ontario, Canada
MOHAMED HEFEEDA, School of Computing Science, Simon Fraser University, Burnaby, British Columbia, Canada

This article presents a review on the process of estimating player engagement in video gaming. To stay ahead of their competitors in entertainment, game developers need to understand, estimate, and maximize player engagement. We address the multidimensional nature of engagement, encompassing cognitive, emotional, and behavioral aspects across various gaming domains. We present a taxonomy of the diverse modalities for quantifying engagement, including physiological signals, observable behaviors, and gameplay data. We identify the challenges of conducting representative subjective studies in this domain and summarize various methods for establishing ground truth measurements. By synthesizing existing research, we provide insights into modeling techniques, highlight research gaps, and offer practical guidelines for implementing engagement measurement strategies. This review aims to aid researchers and industry professionals in navigating the complexities of player engagement estimation, ultimately contributing to enhanced game design, marketing, and user retention in the competitive gaming landscape.

CCS Concepts: • **Human-centered computing** → **HCI design and evaluation methods**; • **Applied computing** → **Computer games**; • **Computing methodologies** → *Machine learning*;

Additional Key Words and Phrases: Player Engagement, Multimodal Data, Engagement Estimation, Video Game Analytics

ACM Reference format:
Ammar Rashed, Shervin Shirmohammadi, Ihab Amer, and Mohamed Hefeeda. 2025. A Review of Player Engagement Estimation in Video Games: Challenges and Opportunities. *ACM Trans. Multimedia Comput. Commun. Appl.* 21, 7, Article 192 (July 2025), 33 pages.
https://doi.org/10.1145/3722116

1 Introduction

The video gaming industry now generates more revenue than the music and movie industries combined [4]. Video gaming is a type of entertainment, and gaming companies strive to "out-fun"

This work is supported by the Natural Sciences and Engineering Research Council of Canada (Grant ALLRP556311-20).
Authors' Contact Information: Ammar Rashed (corresponding author), School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Ontario, Canada; e-mail: arasi005@uottawa.ca; Shervin Shirmohammadi, School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Ontario, Canada; e-mail: shervin.s@uottawa.ca; Ihab Amer, Advanced Micro Devices Inc, Markham, Ontario, Canada; e-mail: ihab.amer@amd.com; Mohamed Hefeeda, School of Computing Science, Simon Fraser University, Burnaby, British Columbia, Canada; e-mail: mhefeeda@sfu.ca.

ACM 1551-6865/2025/7-ART192
https://doi.org/10.1145/3722116

competitors and attract more customers. As such, maximizing player engagement has become crucial. To maximize it, game developers need to first gauge players' engagements over parts or entire gameplays. While players can be stopped, in the middle of the game, and asked about their engagement level, this approach is both intrusive and unscalable. Automatic estimation, if accurate enough, would be more practical. However, as games grow more extensive and complex, the automatic estimation of player engagement becomes increasingly challenging. This article reviews methods for modeling and estimating player engagement automatically from multimodal data, including physiological signals such as **Heart Rate (HR)**, respiration, and skin conductance, neurological signals such as **Electroencephalogram (EEG)**, facial signals such as facial expressions, eye data, and head movements, eye metrics such as pupil size, blink rate, and gaze movements, and gameplay features such as game telemetry, pixel data, user inputs, player skill, and game difficulty. Our study aims to be a one-stop shop aiding researchers and developers in navigating the complex aspects of player engagement estimation.

The gaming industry has been proactively developing tools and strategies for engagement measurement and enhancement. Recent industry reports indicate several key trends: a shift towards immersive gaming experiences, with nearly half of the top 30 games being immersive [21], widespread adoption of live-service models showing 50% expansion [2], integration of AI for dynamic player interactions [81], and implementation of cross-platform experiences to maintain consistent engagement [102]. Companies are increasingly employing data-driven personalization [48] and collaborating with content creators [92] to enhance player engagement. The industry commonly tracks metrics such as **Daily Active Users (DAU)**, **Monthly Active Users (MAU)**, retention rates, session length, and player churn to quantify engagement [44, 58]. Analytics platforms offer sophisticated tools for tracking event-based player retention and custom engagement triggers [45], enabling developers to make data-driven decisions about game design and player experience. This parallel development in industry and academia suggests opportunities for collaborative research to combine practical applications with theoretical frameworks.

Player engagement is a multidimensional concept involving cognitive, emotional, and behavioral aspects of gameplay [3]. It interrelates with the player's motivation to start or continue playing, influenced by factors like peer pressure, promotional materials, and franchise affinity. Engagement has been studied through various social and psychological theories, each defining it in terms of aspects like skill-challenge balance, cognitive load, or game world immersion [28]. This complexity complicates establishing a standard definition and scale for engagement. This review article focuses on works that quantify, estimate, predict, or measure engagement rather than delving into its theoretical, psychological, and social aspects or qualitative analysis.

Another complexity in estimating player engagement is the variety of gaming domains and platforms. AR and VR applications emphasize immersion and presence (defined in the next section) [8, 101]. Mobile games, often played through touch events and in flexible settings, introduce unique engagement elements, such as advertisement segments, distinct from PC or console gaming [46]. Serious or educational games focus on improving retention and learning performance, offering a different perspective on engagement compared to games focused solely on fun. This review is limited to entertainment video games, with other domains reserved for future work.

The multidimensionality of player engagement has led to various modalities for its estimation. Physiological signals include HR variability, blood pressure, and **Electrodermal Activity (EDA)** [36]. EEG signals have been central in many studies, where an *engagement index* is often designed based on predefined EEG frequency bands and channels [94]. While rich in data, these physiological signals are noisy due to encoding information from various body functions, making accurate engagement estimation challenging. Other research has focused on observable signals from the head, face, and eyes. Head pose and movement can indicate posture changes, like leaning forward or backward, reflecting different emotional states [10]. Facial features have been used to model engagement, either through deep learning models [87] or high-level **Facial Action Units (FAUs)** [39]. Additionally, eye-tracking studies have estimated engagement levels by analyzing gaze movement, pupil dilation, and blinking rates [115]. Extending beyond the player's body, gameplay pixel data and user inputs (e.g., keystrokes) have also been used to estimate engagement [83]. This article summarizes the various features used to estimate player engagement and provides a taxonomy for their modalities.

Finally, mapping these features to an explainable scale of player engagement is complex and requires accurate ground truth labels for analysis and validation. The subjective nature of player engagement complicates estimation, often relying on self-reports or observer annotations. Players might complete questionnaires about their emotional and cognitive state, such as the Game Experience Questionnaire [47], or annotate gameplay recordings with perceived engagement levels [83]. Interviews and focus groups offer additional subjective assessments. To address the complexity of subjective engagement, some researchers simplify and operationalize engagement into quantifiable concepts, such as the desire to continue playing, i.e., conation [89], or by manually selecting segments of self-recorded gameplay [20]. Game levels can also be manipulated to include engaging design elements, stimulating different engagement levels in players [75]. This article reviews the various approaches to measuring player engagement, discussing their advantages and limitations, especially in the context of validating engagement estimation models.

In summary, in this article we establish the conceptual foundations of player engagement and review its applications in the video gaming industry. Our main focus is on taxonomizing the various modalities for estimating player engagement, ground truth measurement methods, and modeling approaches. We then discuss the pros and cons of these methods, aiming to identify research gaps and offer practical insights for selecting and implementing effective player engagement measurement strategies. This, in turn, can enhance game design, marketing, and customer retention.

Figure 1 shows the overall framework of player engagement estimation, which also constitutes the roadmap for this article. The entire figure illustrates the needed activities for building an engagement estimation model, while the dashed box shows the activities that are also performed during inference time. First, engagement has to conceptually defined, as covered in Section 2, where we explain the necessary background and technical terms, and propose a definition of engagement that captures its common essence. Next, various predictors of engagement must be taken into account, as covered in Section 3, where we present and taxonomize the predictors used in the existing literature. This is followed by determining engagement's ground truth, covered in Section 4, where we present existing methods and their comparison. The ground truth is required for the eventual validation of engagement estimation methods, which are covered in detail in Section 5. Finally, we end the article in Section 6 by summaries, conclusions, and future work.

2 Background and Definitions

Engagement is a multidimensional construct encompassing cognitive, emotional, and behavioral aspects. Its broad and ambiguous nature makes it difficult to define, model, or quantify. Given the variety of constructs associated with player engagement, it is best studied through a combination of more concrete concepts. Engagement is typically defined using complementary terms that collectively form a mosaic-like model. This section reviews common definitions related to player engagement in games and proposes a comprehensive definition capturing the common essence of engagement.

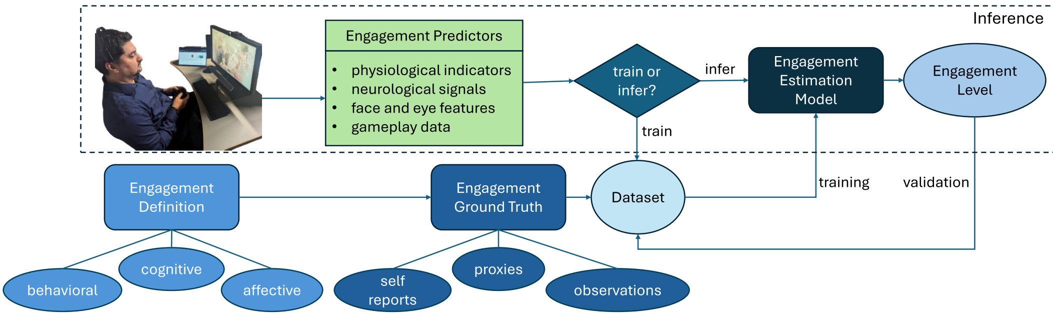


Fig. 1. General framework for player engagement estimation in video games.

2.1 Definitions

2.1.1 Player Modeling. Player modeling studies the player's experience and the dynamic phenomena of gameplay interaction [120]. It involves measuring and representing a player's skill level as a data structure [90]. While primarily based on game-player interaction dynamics, player modeling can also include static player profile information, encompassing cognitive, affective, behavioral, and demographic aspects.

Player modeling approaches can be top-down (model-based) or bottom-up (model-free). Top-down uses theoretical frameworks from social sciences [119]. For instance, emotional models (e.g., valence and arousal [34]) can scale emotional states, while cognitive and behavioral models, such as usability theory [50], aid game design, and refinement. However, these theories often lack empirical validation in gaming contexts. Bottom-up (model-free), by contrast, fits models to player data without strong assumptions about the model. These include predicting player actions, detecting behavioral patterns, or identifying player states [29, 106]. Model-free approaches bridge abstract psychological constructs with quantifiable observations, as seen in psychophysiology studies where facial expressions, head poses, and physiological signals (e.g., HR, blood pressure) are used to gauge player engagement [71, 104]. Hybrid approaches integrate model-based theories with machine learning to map player data to latent states [120].

2.1.2 Affect. Affect refers to aspects of gameplay that describe the player's emotional state. The Pleasure-Arousal-Dominance model defines affect through three dimensions [67]: "valence," which encompasses the spectrum of emotions, e.g., pleasure to sadness; "arousal," which reflects the intensity of emotion; and "dominance," which relates to the level of control over an emotion, referred to as "autonomy" in other contexts.

2.1.3 Motivation. Motivation is defined as a trait-like personal orientation towards a task [112]. It is often captured by involvement, reflecting the player's perceived relevance to a goal. Motivation can be intrinsic or extrinsic [95]. Intrinsic motivation involves engaging in an activity for its inherent satisfaction, such as enjoying a game that leaves a pleasant memory and increases the desire to play again [78, 95]. Games that offer goals, such as improving skills or completing tasks, enhance intrinsic motivation. Video game designers aim to provide "novelty" to resonate with innate curiosity, a typical intrinsic motivator. In contrast, extrinsic motivation involves engaging in an activity for a separate outcome [95]. This can range from reluctant compliance, such as studying to avoid sanctions, to self-endorsed goals, like career opportunities or social status. Ideally, games should cater to both types of motivation: offering enjoyable experiences for intrinsic motivation and rewards, incentives, or social interactions for extrinsic motivation. For instance, in multiplayer games, players might join a "clan" and contribute regularly, even if the tasks are not inherently fun.

2.1.4 Immersion. Immersion is defined as a psychological state where one feels enveloped by, included in, and interacting with an environment that continuously provides stimuli and experiences [116]. Often conflated with presence, immersion relates to the player's perceived presence in a virtual environment [56]. Early research by [15] identified immersion, engagement, and engrossment as key factors of a player's experience. Ermi and Mäyrä [32] categorized immersion into Sensory, Challenge-based, and Imaginative types. Sensory immersion depends on the game's audiovisual execution, Challenge-based immersion relates to satisfying challenges, and Imaginative immersion involves role-playing elements like story and characters. Different game genres stimulate varying levels of these immersion types. Unlike player engagement, immersion does not directly address behavioral aspects. Jennett et al. [51] defined immersion through flow, presence, and cognitive-absorption. Flow, similar to challenge-based immersion, involves a balance between challenge and skill. Cognitive-absorption and presence correspond to sensory and imaginative immersion, respectively. The construct of immersion in video games has been summarized as follows:

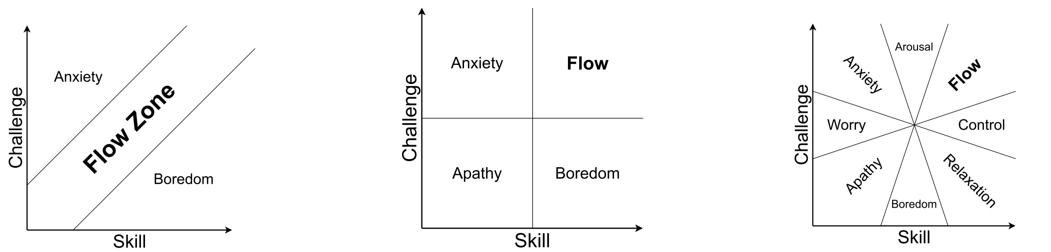
> "Immersion is a subjective state characterized by perceiving oneself to be enveloped by, included in, and interacting with a video game that provides a continuous stream of stimuli and experiences. Immersion requires focused attention on a limited stimulus field and minimized distractions which can be promoted by the video game system itself. Immersion may be enhanced by the capability of the video game's technology to provide the player immersive cues. This includes the ability to interact with the video game through a virtual representation of the player. Interaction must seem natural with regard to the input mechanisms and the game's response to the player. Immersive cues are also strengthened by increasing the extent, fidelity and resolution of sensory information. Lacking immersive cues, involvement and individual differences may mitigate the deficit, thus helping the player to experience immersion." [86]

2.1.5 Presence. The concept of presence originally refers to the phenomenon of *externalization*, in which one's perception is referenced to an external space beyond the limits of the sensory organs [66]. In the context of video games, the space is the game environment where the player's natural perception is focused. Presence is affected by various elements in game design, such as display, co-player, and co-playing modes [17]. The concept of presence has been defined as follows:

> "Presence is a state of conviction of being located in the game environment. It is a binary experience, during which perceived self-location and perceived action possibilities are connected to the game environment, and mental capacities are bound by the game environment instead of reality." [86]

2.1.6 Engrossment. Engrossment involves *emotional attachment* and *decreased perception* as described in [3], which performed interviews with participants about engaging experiences and concluded that engrossment affects awareness of surroundings. One interviewee noted, "I didn't notice really that it was getting darker," indicating a loss of spatiotemporal awareness. Conversely, another gamer described engrossment as choosing to focus solely on the game, stating, "you have to almost seclude yourself because this is what I have to do for the next couple of hours." In video games, graphics quality and visual realism significantly impact engrossment [13], which explains why game designers aim for visually stunning graphics to enhance player engrossment [114].

2.1.7 Involvement. Involvement encapsulates the cognitive and motivational aspects of gameplay experiences [86]. Although distinct from immersion, involvement is reciprocally related to it as one increases the other. Generally, involvement is defined as "person's perceived relevance of the object based on inherent needs, values, and interests" [121]. In the context of video games, involvement has been defined as follows:



(a) Basic Model. (b) Quadrant Model. (c) Experience Fluctuation Model.

Fig. 2. Models of flow.

> "Involvement is a motivational factor regarding gameplay that is experienced as a sequence of focusing one's energy and attention on a coherent set of stimuli or meaningfully related set of activities or events. Involvement depends on the degree of perceived relevance that the individual attaches to the stimuli, activities, or events. Involvement is increased in by playing video games that stimulate, challenge, and engage the user either cognitively, physically, or emotionally. Involvement has a reciprocal relationship with immersion, where increasing a sense of immersion similarly increases a sense of involvement, and vice-versa." [86]

2.1.8 Flow. Flow, akin to immersion, involves cognitive absorption where players lose awareness of their surroundings. Described as the "optimal experience" [24], flow occurs when a game presents a satisfying challenge relative to the player's skill level. Procci [86] identifies nine elements of flow: (1) challenge/skill balance; (2) concentration; (3) clear goals; (4) immediate feedback; (5) merging of action and awareness; (6) control; (7) loss of self-consciousness; (8) time distortion; and (9) the autotelic experience. This aligns with earlier focus on challenge/skill balance, immersion, control, and other factors [18, 22]. Flow theory [24] emphasizes balancing challenge and skill, with anxiety and boredom representing extremes of this balance. A well-designed game ideally transports players to their "Flow Zones," which monopolize attention, requiring sustained focus and reducing awareness of distractions [78], and fostering pleasure and happiness [19]. Flow's relevance to gameplay analysis lies in its focus on challenging players to reach their skill limits [32, 112].

Flow theory models engagement as a function of player skill and game challenge. The most basic model indicates that players experience high anxiety when challenges greatly exceed skills and boredom when the game is too easy, with the flow zone being the balance between these extremes (Figure 2(a)). The Quadrant model [74] (Figure 2(b)) builds on this by categorizing player states into apathy, boredom, anxiety, and flow based on skill and challenge levels—flow occurs when both are high, and apathy when both are low. The Experience Fluctuation model [6] (Figure 2(c)) further refines this by dividing player states into eight dimensions of experience, accounting for mid-levels of skill and challenge.

2.1.9 Endurability. Endurability is defined as "the likelihood of remembering enjoyable situations and intending to perform them again" [78]. It involves two aspects: the Pollyanna principle, which suggests a tendency to remember pleasant experiences more than unpleasant ones [26], and "returnance," where enjoyable experiences increase the desire to repeat them [88]. Endurability reflects how a well-designed game can boost a player's motivation to return, impacting gaming behavior. It complements immersion and flow by focusing on the behavioral effects of enjoyable experiences.

Table 1. Qualitative Comparison of Engagement-Related Constructs

Construct	Cognitive Aspect	Affective Aspect	Behavioral Aspect
Motivation	Perceived relevance to game. Curiosity. Interest in game theme/genre.	Pleasure-driven incentive.	Agency. Action management. Goal-oriented.
Immersion	Audiovisual cues for sensory immersion. Spatiotemporal distortion. Concentration.	Imaginative immersion in the game world and story. Emotional attachment. Reciprocal with fun.	Volitional seclusion.
Presence	Externalization: envelopment in the game's virtual environment.	Reciprocal with emotional attachment to game world and story.	Commitment to play the game.
Engrossment	Cognitive absorption.	Emotional attachment to the game experience.	Commitment to play the game.
Involvement	Perceived relevance to player needs, values, and interests. Cognitive absorption focusing on game objectives. Reciprocal with immersion.	Reciprocal with fun.	Encapsulates motivation.
Flow	Max cognitive absorption. Challenge-skill balance.	Optimal experience. Emotional attachment. Lingering positive memories.	Clear goals. Immediate feedback. Encapsulates immersion and motivation.
Endurability	Reciprocal with immersion.	Requires strong enjoyment.	Returnance and replayability. Long commitment to the game.

2.2 Taxonomy and Lessons Learned

The taxonomy in Table 1 reveals several key insights. First, engagement emerges as a dynamic process rather than a static state, incorporating multiple complementary constructs. Second, while each construct exhibits unique characteristics, they share overlapping cognitive, affective, and behavioral dimensions. Third, these constructs form a progressive chain: motivation initiates engagement, immersion deepens it, flow optimizes it, and endurability sustains it.

Consider a player's progression through this engagement chain: initially *motivated* by social recommendations or intrinsic interest, they begin playing. Audiovisual elements capture attention, evolving into deep concentration and environmental unawareness, indicating cognitive absorption. This fosters emotional attachment to game elements, leading to *immersion* and subsequently *engrossment*. The player's motivation and cognitive absorption reflect their *involvement*. With clear goals and feedback, optimal challenge-skill balance leads to *flow*, providing peak engagement. The resulting satisfaction contributes to *endurability*, reinforcing return motivation. Figure 3 illustrates this model.

This understanding yields two critical implications: (1) engagement cannot be reduced to a single construct but must be understood as their collective interaction, and (2) effective engagement measurement requires a multimodal approach capturing these various dimensions and their temporal progression.

3 Player Engagement Predictors

In this section, we describe and categorize various signals used to quantify player engagement. Player engagement can be measured through a diverse array of modalities:

— *Physiological* indicators include HR, respiration, and skin conductance.
— *Neurological* signals, mainly EEG.

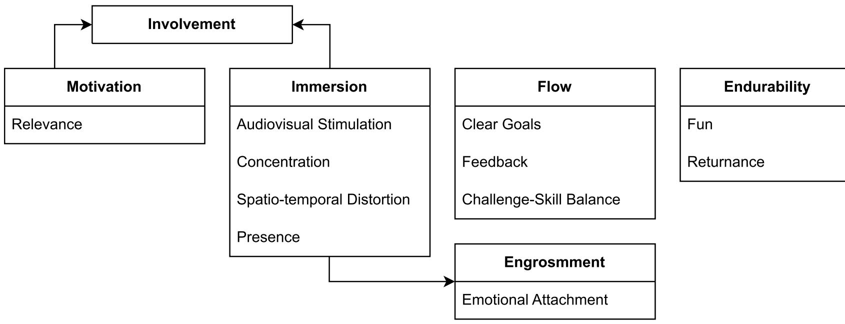


Fig. 3. An illustration of player engagement composition as a process (left to right).

— *Facial* signals include facial expressions, eye data, and head movements.
— *Eye* metrics include pupil size, blink rate, and gaze movements.
— *Gameplay* features include game telemetry, pixel data, user inputs, player skill, and game difficulty.

By organizing these signals into broader categories, we aim to present a comprehensive taxonomy that facilitates the analysis and interpretation of engagement data within the context of game design and player experience.

3.1 Modalities

3.1.1 Physiological Signals. Physiological signals provide insights into player engagement by recording automatic bodily responses to game stimuli. These signals can be categorized by their relation to systems or body aspects, including cardiovascular, respiratory, EDA, temperature, and muscle responses, based on traditional psychophysiology [33]. These are described next.

Cardiovascular. Cardiovascular signals include HR and **Heart Rate Variability (HRV)**. HR can be measured using **Electrocardiogram (ECG)** or **Photoplethysmography (PPG)** sensors. While ECG is generally more accurate, PPG is common in wearables like smartwatches and provides comparable accuracy [110]. HRV metrics, such as the root mean square of successive differences and mean standard deviation of RR intervals, are derived from inter-beat intervals and are used to detect arousal and stress, emotional aspects of player engagement [7, 63]. HRV is also relevant for identifying video game addiction [55]. Blood pressure, though less popular, can also be used for similar purposes [18]. The use of HRV in esports has been reviewed in [111], noting that while promising, it often lacks a solid theoretical foundation and robust methodology.

Respiration. Respiration rate is measured using belts that track chest cavity expansion, with features including inspiration/expiration time, apnea, and respiration interval [107]. While some studies find no correlation between respiratory features and self-reported enjoyment [107], others use it to predict player fun levels [35]. Respiration intensity can also be measured with a digital thermometer placed under the nose, quantifying valence and arousal [38]. However, respiration features are typically used alongside other features in emotion detection, with a lack of robust ablation studies on their importance in emotion estimation models.

EDA. **Galvanic Skin Response (GSR)**, or EDA, measures skin conductance, which varies with sweat gland activity controlled by the sympathetic nervous system. This response reflects emotional and cognitive processing, including reactions to threat, anticipation, and novelty [84]. EDA is useful for assessing stress, with peak height and rate serving as indicators [84]. In video games, EDA data, comprising of tonic (slowly changing) and phasic (event-related) conductance, is analyzed to gauge player engagement and cognitive load [85]. The EDA signal is sampled (e.g., at 15 Hz) and processed using techniques like Butterworth low-pass filtering to separate tonic and phasic components and detect peaks based on skin conductance response rates and amplitudes. Increased sweating, measured by GSR, correlates with higher levels of fun [36]. While EDA and other physiological measures can predict cognitive load, they only explain part of the variance, suggesting other influencing factors [85]. EDA features are more important in fun estimation models than features like pupil diameter, age, and perceived difficulty [35].

Temperature. Skin temperature is used to assess player engagement and emotional states, with changes reflecting different emotional responses during gameplay [18]. It is measured by recording palmar temperature, which indicates autonomic or parasympathetic nervous system activation, often related to cognitive load [85]. Key features include the mean temperature and its average derivative, analyzed using statistical methods like ANOVA to detect differences under various conditions. For instance, increased game difficulty correlates with decreased temperature, suggesting a shift from engagement to boredom with higher skill levels [18]. Correlations between temperature and subjective cognitive load measures, such as frustration and NASA-TLX scale items, show medium-strength relationships [85]. However, the use of skin temperature data can be affected by confounding factors like prior activity and interactions with other physiological signals such as EDA [85]. Despite this, skin temperature remains a valuable measure for understanding player engagement and physiological responses to gameplay.

Electromyography (EMG). EMG measures electrical activity produced by muscles and is valuable for assessing emotional states and cognitive processes in video games [35, 42, 85, 118]. Facial EMG sensors on muscles like the corrugator (frowning) and zygomaticus (smiling) capture changes in muscle tension related to emotional valence and mental effort [42]. EMG can detect emotional responses even without overt facial expressions [42, 118]. Other muscles, such as the biceps brachii, can be monitored for effort and cognitive load [85]. EMG data processing involves cleaning, filtering, and feature extraction to quantify muscle activation [85]. The amplitude of EMG signals correlates with cognitive load, demonstrating its relevance for understanding mental workload [85]. Additionally, EMG sensors integrated into machine learning models can predict game experience aspects like difficulty and immersion, enhancing predictive performance when combined with other modalities [118].

3.1.2 Neurological Signals. EEG is a technique that measures electrical brain activity using scalp electrodes. It offers high temporal resolution, capturing rapid changes in brain activity linked to cognitive and emotional states. Recent advancements, as explored in studies like [9, 11], have enabled real-time assessment of player states and engagement during gameplay.

EEG Measurement. EEG signals are processed to extract features such as power spectral density across different frequency bands (delta, theta, alpha, beta, gamma). Each band reflects distinct neural processes [80]:

- *Delta (0–4 Hz)*: Delta waves are prominent during deep sleep in adults, characterized by high amplitude and slow frequencies.
- *Theta (4–7 Hz)*: Theta waves are associated with relaxed and meditative states, and their synchronization patterns modulate during changes in affective states.
- *Alpha (8–12 Hz)*: Alpha waves are observed during relaxed states and tend to diminish with cognitive exertion. They are linked to both negative and positive valence states and frontal asymmetry.
- *Beta (12–30 Hz)*: Beta waves, of low amplitude, are prevalent during cognitive processes such as thinking and concentration.

Table 2. EEG Frequency Bands

Band	Bandwidth	Neural Process
Delta	0–4 Hz	Sleep and Dreaming
Theta	4–8 Hz	Deep Relaxing and Meditation
Alpha	8–12 Hz	Resting and Relaxing
Beta	12–30 Hz	Alert and Active Mind
Gamma	>30 Hz	Intense Focus and Problem Solving

— *Gamma (>30 Hz)*: Gamma rhythms are associated with the binding of neural networks performing specific cognitive functions, reflect changes in affective states, and are influenced by stimuli like aversive visual cues.

Table 2 summarizes common EEG frequency bands. It is worth noting that the frequency bandwidth of these bands can differ slightly in different works.

EEG Engagement Indices. Several EEG-based indices have been developed to quantify player engagement during gaming activities. These indices typically focus on specific frequency bands and their ratios, reflecting different aspects of cognitive processing and emotional arousal. The most important of these indices are:

— *Beta/(Alpha + Theta)*: This index has been shown to correlate with varying levels of cognitive load and arousal during gameplay [53].
— *Frontal Theta*: This index focuses solely on the frontal theta band activity, which is associated with cognitive engagement and effortful processing. It has been shown to have relevance in distinguishing between different gaming modalities and task demands [37].
— *Frontal Theta/Parietal Alpha*: This index is particularly sensitive to changes in cognitive workload and attentional processes during gaming tasks [72].
— *Theta AF3/Alpha P7*: Proposed in [94], this index uses the ratio of theta band activity at the AF3 electrode to alpha band activity at the P7 electrode. It aims to capture specific cognitive engagement patterns during gameplay, distinguishing between different levels of player involvement.

ML Models Using EEG. Using EEG data to estimate player engagement during gameplay, various classifiers have been employed, such as **Support Vector Machines (SVM)**, **Naive Bayes (NB)**, and **k-Nearest Neighbors (kNN)**. For instance, [80] found that the NB classifier was most robust for identifying negative events like character deaths, while kNN, particularly using the Beta band, was better for general gameplay events, suggesting that combining classifiers can be more effective than using a single one. The work in [9] demonstrated that SVM classifiers can classify three levels of user states with reasonable accuracy, with user-dependent classification outperforming user-independent classification (66.4% vs. 50.1%). Hafeez et al. [40] showed that EEG data could classify expert and novice players with up to 98.33% accuracy using kNN. Recent studies also used genetic algorithms for feature selection and clustering, showing effectiveness in identifying EEG patterns related to player involvement [91].

These findings highlight the potential of machine learning in capturing complex EEG patterns for real-time player engagement estimation. However, model success depends on EEG data quality, feature selection, and classifier choice. Consumer-grade EEG devices may have lower spatial resolution and signal quality compared to medical-grade equipment, affecting accuracy and reliability [80]. Additionally, EEG headset comfort and placement can influence user behavior. Despite promising results, further validation against subjective measures and performance metrics is needed, and deep learning models for EEG-based engagement estimation are scarce, suggesting an area for future research.

Commercial Solutions. In addition to academic research, commercial EEG solutions like EMOTIV [31] offer proprietary algorithms for measuring engagement-related metrics, reporting high accuracy in distinguishing engagement levels in controlled settings. However, as these solutions are proprietary, their methodologies are not fully available for academic validation or reproduction.

3.1.3 Facial. Facial signals can be crucial for understanding player engagement, offering detailed insights into emotional and cognitive states. They provide information that traditional metrics may miss [87]. Using devices like webcams, facial signals can be recorded unobtrusively and analyzed in real-time during gameplay. Advances in computer vision have improved the accuracy of detecting subtle emotional and behavioral cues [20]. Key facial signals include:

- FAUs, which capture muscle movements linked to emotions;
- Facial embeddings, used for identifying emotional patterns [20];
- Facial expressions, recognized through emotion recognition algorithms [54];
- Head movements, reflecting interest and immersion [98].

These signals provide a comprehensive toolkit for researchers and developers to enhance user experience and tailor game dynamics to player preferences. Next, we describe each of them.

FAUs. FAUs capture specific facial muscle movements that indicate emotional and cognitive states, crucial for quantifying player engagement in video games. For example, FAU12 (lip corner puller) and FAU6 (cheek raiser) often signal joy or satisfaction [39]. Detected using facial landmark detection and tracking algorithms, FAUs analyze changes in facial geometry and muscle activations in real-time [20]. Studies show that FAUs effectively identify engagement levels and emotional responses during gameplay, highlighting their utility across various game genres and skill levels [87]. They enable precise monitoring of player reactions and adaptation of game dynamics, offering interpretable insights into emotional and cognitive states [64].

Facial Landmarks and Embeddings. Facial landmarks are specific points on the face, such as the corners of the eyes and mouth, the nose tip, and the eyebrows, used to track facial expressions and movements in real-time [20]. Facial embeddings, on the other hand, are dense numerical representations that capture unique facial patterns for identification and emotion recognition [108]. EmoNet, for example, uses deep neural networks to estimate emotional dimensions like valence and arousal from facial images, improving accuracy under natural conditions [108].

Facial Expressions. In player engagement assessment, facial expressions reveal emotions such as happiness, sadness, anger, and surprise through patterns of muscle movements [87]. While FAUs indicate specific muscle actions, facial expressions provide a broader view of emotional states [39, 113]. This analysis helps adapt game content in real-time to enhance player immersion [49]. Facial embeddings further personalize interactions by responding to individual emotional cues, optimizing gameplay based on the player's affective state [113].

Head Pose. Head pose features, which capture the orientation and positioning of a player's head during gameplay, can be used to assess player engagement. These features are extracted using computer vision techniques that analyze head angles and movements in real-time. Head pose data offers insights into a player's attention, focus, and emotional state based on their spatial interactions within the game environment [98]. Key head pose features include:

- *Yaw, Pitch, and Roll*: These are fundamental angles that describe the orientation of the head relative to a fixed reference frame. Yaw refers to rotation around the vertical axis (left-right movement), pitch around the lateral axis (up-down movement), and roll around the longitudinal axis (tilt side-to-side) [109].

- *Head Movement Dynamics*: This encompasses the speed, frequency, and smoothness of head movements during gameplay. Rapid or frequent head movements may indicate heightened engagement or reaction to game stimuli, whereas minimal movement could suggest disengagement or distraction [52].
- *Gaze Direction*: Although primarily associated with eye tracking, head pose can provide an indirect measure of where a player is looking within the game interface. Changes in head orientation relative to the screen can infer shifts in visual attention and cognitive processing [49].
- *Head Alignment*: Analyzing the alignment of the head with respect to specific game elements or events can reveal patterns of interest, such as focusing on opponents or exploring new environments [109].

Head pose features are crucial for understanding player behaviors and cognitive responses during gameplay, offering valuable insights into engagement and immersion dynamics in video games [98]. By integrating analyses of head orientation, movement dynamics, and gaze direction, game developers can tailor content in real-time to optimize challenge levels and enhance player experiences [52, 109]. These advancements underscore the growing importance of head pose recognition technologies in both research and practical applications within the gaming industry, where they contribute to refining player engagement models and improving overall game design strategies [49].

3.1.4 Eye. In the study of player engagement in video games, several key eye metrics have been employed to understand players' visual attention and emotional states:

- *Pupil Size*: Pupil dilation is a key indicator of arousal and cognitive load during gameplay [60]. Larger pupil size typically reflects increased arousal, which can indicate either heightened engagement or stress, depending on gameplay challenges [52]. For example, Lu et al. demonstrated that pupil size variations can reveal different levels of engagement and cognitive processing during educational game tasks, highlighting its effectiveness in assessing player involvement and cognitive effort [60].
- *Blink Rate*: Blink rate is used to infer cognitive workload and affective states like frustration and confusion during gameplay [115]. A higher blink rate often signals increased cognitive load and stress, especially in challenging game scenarios where players may experience frustration [115]. By analyzing blink rate, researchers can gain insights into how players emotionally and cognitively respond to game challenges, shedding light on their engagement levels and emotional responses [52].
- *Fixations and Saccades*: Fixations and saccades are crucial for understanding visual attention and strategy deployment in games [16]. A lower fixation/saccade ratio typically suggests more exploratory behavior or difficulty in processing game elements, which can vary with game complexity and player skill levels [115]. For instance, Burch et al. explored how these metrics reveal players' visual attention patterns across different game modes and scenarios, offering insights into effective game design and player engagement strategies [16]. Figure 4 illustrates the relationship between fixation and saccade.
- *Relevancy Ratio*: The relevancy ratio measures the proportion of fixations on relevant vs. irrelevant game elements, offering insights into players' strategic focus and decision-making [115]. A higher relevancy ratio indicates effective attention direction towards game-relevant cues, correlating with greater engagement and task efficiency [76]. Ninaus et al. showed

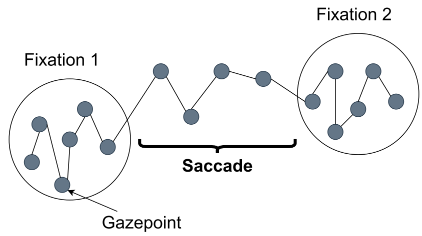


Fig. 4. The relationship among gaze point, fixation, and saccade.

that this metric can distinguish between engaging and non-engaging tasks, highlighting its importance in optimizing game mechanics to sustain player interest and challenge [76].

Despite their utility, eye metrics have limitations in quantifying player engagement. Their interpretation is context dependent, influenced by factors like game genre, player experience, and cultural differences, which affects their generalizability [52]. Variability in eye-tracking devices, whether high-end or webcam based, can impact the accuracy of metrics like fixation duration and saccade rates, leading to potential measurement errors and difficulties in comparing results across studies [117]. Additionally, interpreting these metrics requires considering factors such as task complexity and cognitive workload, as ambiguities can arise in dynamic and complex gaming environments [115]. Acknowledging these limitations can help refine the use of eye-tracking metrics in gaming research, enhancing our understanding of player engagement and informing game design strategies.

3.1.5 Gameplay. Gameplay features are crucial for quantifying player engagement, offering insights into behavior, preferences, and emotional responses. They can be categorized into:

— *User Inputs*: These include actions such as keystrokes, mouse clicks, or controller movements that reflect player decisions and interactions within the game.
— *Pixel Data*: This encompasses visual information from the game screen, including changes in the player's view, object appearances, and visual effects that contribute to the overall game experience.
— *In-Game Events (Telemetry Data)*: These are data points related to game events and player actions, such as achievements, failures, or interactions with game elements, which help analyze engagement levels.

Utilizing these features effectively enables developers and researchers to model engagement, predict player retention, and enhance game design to keep players interested, as described next.

User Inputs. User inputs refer to the interactions players have with the game through devices such as gamepads, keyboards, or mice. These inputs capture detailed actions, including button presses and joystick movements, offering insights into player behavior. They reflect immediate actions and decisions during gameplay and are crucial for understanding how players engage with the game environment and respond to challenges. For example, in a study [83] of Tom Clancy's The Division 2, detailed gamepad actions were logged and analyzed to predict player engagement. The analysis included 25 different gamepad actions and their co-occurrences, providing insights into interaction patterns. The study found that while user inputs offer detailed behavioral data, they often lack context about the outcomes of these actions within the game, making it challenging to directly infer engagement levels. Additionally, individual playing styles can introduce variability, requiring sophisticated models to extract meaningful patterns from the data.

Pixel Data. Pixel data refers to frames of in-game footage that capture the visual context and the environment in which players interact. This data provides insight into the graphical and visual elements of the game, contributing to the overall gaming experience. For instance, in the said study of Tom Clancy's The Division 2, high-resolution gameplay frames were analyzed to predict long-term player engagement. The visual context from these frames was essential for understanding player behavior and actions within the game [61, 62]. However, processing pixel data is computationally intensive and requires substantial storage capacity. Additionally, while pixel data provides detailed visual information, it may not fully capture players' emotional states or the underlying reasons for their actions. Therefore, combining pixel data with other data types, such as user inputs or physiological signals, is often necessary for a comprehensive understanding of player engagement [61, 62].

In-Game Events. In-game events, or game telemetry data, offer detailed logs of player activities, progression, and interactions within the game. This data encompasses various aspects of player behavior, such as playtime, mission completion, and progression levels, providing insights into the player's journey through the game. For instance, in a study on PUBG streaming on Twitch, 40 gameplay features derived from telemetry data were used to model viewer engagement, leveraging data from hundreds of matches and over 100,000 game events to demonstrate the scalability and effectiveness of telemetry data in predicting engagement [69]. Similarly, in Tom Clancy's The Division, high-level gameplay metrics were correlated with player motivation, capturing aspects like playtime and mission completion [68]. Despite its utility, telemetry data can be coarse-grained, potentially missing the subtleties of player experience and motivation. For example, while telemetry data can indicate mission completion, it may not capture the emotional response or challenges faced by the player during that mission. Additionally, telemetry data is often game specific, which can limit its generalizability across different genres and titles [68, 69].

User Outputs. User outputs provide unique insights into player engagement through **User-Generated Content (UGC)** and social interactions. UGC, including custom roles, levels, and maps, demonstrates deep engagement through creative investment in the game [30]. Social communications between players, particularly in multiplayer settings, offer rich engagement indicators through chat frequency, sentiment, and collaborative patterns [5, 82]. Viewer engagement in streaming contexts can also serve as a proxy for gameplay engagement, with chat activity correlating with engaging gameplay moments [69]. However, these measures face challenges including privacy concerns, data accessibility, and the need for sophisticated natural language processing to interpret communication content.

Industry Perspective. In industry practice, gameplay features are tracked through comprehensive analytics platforms that monitor various engagement KPIs [44]. These include:

- *Session Metrics*: Length and frequency of gameplay sessions, indicating immediate engagement;
- *Retention Metrics*: Player return rates at various intervals (1-day, 7-day, 30-day);
- *Monetization Indicators*: Conversion rates, average revenue per user, and lifetime value;
- *User Acquisition Metrics*: Install rates, new user acquisition, and associated costs.

These industry metrics complement academic research by providing practical insights into player engagement patterns and their business impact [1, 58].

3.2 Taxonomy and Lessons Learned

The comprehensive review of engagement predictors reveals several critical insights. First, each modality offers unique perspectives: Physiological signals provide objective measures of arousal

Table 3. Qualitative Comparison of Player Engagement Predictors

Data Source	Features	Papers	Engagement Aspects	Practical Aspects
Physiological Signals	Cardiovascular Respiration EDA Temperature EMG	[18, 35] [38, 42] [55, 73] [84, 85] [107, 111] [118]	Game Addiction Cognitive Load Fun Excitement Stress	Measured through invasive devices. Estimatable through non-invasive widgets (e.g., smartwatches or cameras). Sensitive data raises privacy concerns. Requires stronger theoretical foundation and robust methodology.
EEG	Beta/$(\alpha + \theta)$ Frontal θ Frontal θ/Parietal α θ AF3/α P7	[9, 11, 37] [40, 53] [72, 80] [91, 94]	Cognitive Load Arousal Attention Patterns Involvement	Requires special invasive devices. Very sensitive to noise. Processing signals from several channels is computationally expensive. Generalizability of engagement indices requires further research.
Face	FAUs Facial Landmarks Head Pose	[20, 108] [39, 49] [52, 54] [87, 98] [109, 113]	Emotional and Cognitive States	Accessible via webcams. Widely shared by game streamers. Pre-trained models for feature extraction are common.
Eye	Pupil Size Blink Rate Fixations and Saccades Relevancy Ratio	[16, 52] [60, 76] [115, 117]	Cognitive Load Stress Attention Patterns	Measured with special eye-trackers. Partially estimatable from face. Sensitive to specific game contexts and individual player characteristics.
Gameplay	User Inputs Pixel Data In-Game Events User Outputs	[61, 62] [68, 69] [83] [5, 30, 82]	Interaction Patterns Visual Stimuli Motivation Creative Investment Social Engagement	Noise due to individual playing styles. Cannot capture emotional responses. Cannot directly be generalized between different games or genres. Privacy concerns and data accessibility. May require language processing.

and cognitive load, EEG captures fine-grained neural responses, facial signals offer unobtrusive emotional state detection, eye metrics reveal attention patterns, and gameplay features reflect behavioral engagement. Second, while each modality has distinct advantages, they also face specific limitations: Physiological and EEG signals require specialized equipment, facial and eye metrics need careful contextual interpretation, and gameplay features lack emotional context and cross-game generalizability.

Most importantly, we observe that engagement prediction benefits from multimodal approaches, as each data source compensates for others' limitations. For instance, gameplay features provide behavioral context for physiological responses, while facial expressions help interpret EEG signals. Table 3 systematically compares these modalities, highlighting their complementary nature in capturing different engagement aspects. This understanding suggests that effective engagement measurement systems should:

— Balance invasiveness with measurement accuracy;
— Consider practical deployment constraints in gaming environments;
— Account for individual differences and gaming contexts;
— Integrate multiple modalities while respecting computational limitations;
— Validate measurements against established engagement constructs.

These insights inform both research directions and practical implementations in game development, emphasizing the need for context-aware, multimodal approaches to engagement estimation. The inclusion of user outputs as engagement predictors, particularly through UGC and social communications, offers promising new directions for capturing deeper, longer-term engagement patterns, though careful consideration must be given to privacy and data processing challenges.

4 Player Engagement Ground Truth

To validate an estimated player engagement level, we must know the ground truth first. In this section, we explore various approaches in establishing the ground truth of player engagement level and validating the estimated level. We discuss several validation methods, including self-report methods like questionnaires and user annotations of gameplay recordings, expert evaluation methods such as observations and interviews, and heuristic and proxy methods involving concepts like conation and specially designed game levels. Each of the following subsections delves into these approaches, respectively, highlighting their application and significance in assessing player engagement.

4.1 Approaches

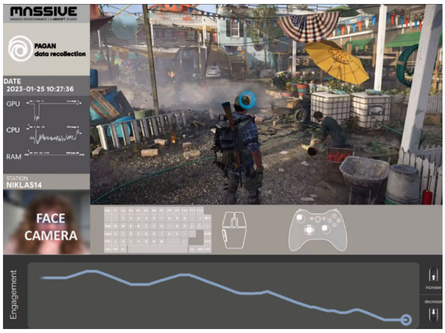
4.1.1 Self-Report Methods. Questionnaires (QREs). Player engagement in video games is often assessed using various questionnaires designed to capture different facets of the player experience. The **Game Engagement Questionnaire (GEQ)** [14] was developed to provide a reliable measure specifically tailored to engagement during video game play. Validated through Rasch analysis, the GEQ is confirmed for its reliability, functional structure, and dimensionality, making it suitable for assessing engagement in gaming contexts.

While the GEQ focuses on absorption, flow, presence, and immersion, other prominent questionnaires address different aspects of player engagement. The **Immersive Experience Questionnaire (IEQ)** [27] measures cognitive and emotional involvement, real-world dissociation, challenge, and control, while the **Player Experience of Need Satisfaction (PENS)** evaluates competence, autonomy, relatedness, control, and presence/immersion. A summary of other questionnaires measuring player engagement is provided in Table A1 in Appendix A.

Ubisoft's **Perceived Experience Questionnaire (PEQ)** and the PANAS are additional tools for assessing player engagement and emotional states in gaming contexts. PEQ captures perceptions of competence, autonomy, relatedness, and presence, as used in Tom Clancy's The Division to understand motivational factors driving engagement [68]. In contrast, PANAS measures affective states like positive and negative emotions experienced during gameplay, useful in predicting engagement levels, particularly in older adults playing mobile games [71]. It was found that higher game performance, prior mobile game experience, and positive affect were significant predictors of engagement. Environmental disturbances negatively impacted engagement, while the number of gameplay sessions and game type did not significantly affect engagement levels. Interestingly, participants with dementia showed increasing engagement over time, and engagement was higher with Word-Search and Mahjong compared to Bejeweled.

Despite their utility, these instruments face challenges in universality across diverse game genres. For example, some GEQ items like "feeling scared" may not apply universally, and PENS items querying relationships with other players are irrelevant in single-player games, leading to confusion and potentially skewed results [27]. Moreover, the interruptive nature of questionnaires limits their ability to capture experience fluctuations throughout a gaming session, as they are typically administered post session, making them unsuitable for real-time engagement measurement.

Continuous Annotations. While questionnaires are useful for gauging general player dispositions, continuous annotations offer the significant advantage of capturing emotional fluctuations during gameplay, providing a dynamic and detailed understanding of player experience. For instance, the RankTrace annotation tool [59] from the PAGAN platform allows players to continuously annotate their engagement while watching gameplay videos, facilitating the capture of nuanced changes over



(a) Data Collection Snapshot via PAGAN Annotation, reused with permission from [83].

(b) Griffin PowerMate

(c) Gazepoint Biometric

Fig. 5. Continuous annotation example tools and interface.

time [83]. This unbounded annotation approach preserves the relative relationships between data points, making subjective experiences easier to interpret [70]. Moreover, continuous annotation reduces guesswork in absolute scales, ensuring a more accurate representation of players' emotional states.

Several studies have effectively employed annotation tools to collect self-reported engagement data. In a study [83] on "Tom Clancy's The Division 2," participants used the RankTrace tool for continuous engagement annotation, enabling researchers to predict long-term engagement accurately by integrating gameplay footage and controller input. Figure 5(a) presents a snapshot of data collection for The Division 2 game [83]. Annotations are typically conducted using a wheel interface, like the Griffin PowerMate (Figure 5(b)) or Gazepoint Biometrics Kit (Figure 5(c)), facilitating continuous and analog self-reported engagement labels. Similarly, the AGAIN dataset employed RankTrace to gather arousal labels across nine games, collecting over 37 hours of annotated gameplay videos, thereby demonstrating the effectiveness of continuous annotation in capturing diverse emotional responses across various genres [70]. These examples underscore the utility of annotation tools in providing detailed, contextually rich engagement data for advanced player modeling and affective computing research.

Despite their advantages, continuous annotation methods face limitations. A key challenge is recall bias; when players annotate engagement after gameplay, the time gap can lead to inaccuracies as they may struggle to remember and report their feelings from earlier moments [41]. Additionally, while continuous methods like RankTrace reduce some guesswork, they still rely on players' subjective interpretation of their engagement, which can introduce variability in the data. Lastly, the requirement for players to watch their gameplay videos for annotation purposes can be time-consuming, potentially affecting data reliability due to annotation fatigue, even if measures like speeding up video playback are employed [70, 83].

4.1.2 Observational Methods. Observational methods, such as third-person expert annotations and interviews, have also been used for quantifying player engagement in video games. These methods offer insights often missed by automated systems or self-report measures. Third-person annotations add objectivity, as expert annotators—usually experienced players or researchers—apply consistent knowledge to evaluate engagement, reducing the bias from individual players. Moreover, interviews capture nuanced experiences and motivations that players may struggle to express quantitatively, enriching the data collected.

Self-report methods assume players have high self-awareness to accurately report their emotions. However, many studies employ a *third-person* annotation protocol, where expert teams assess player engagement based on observation or interviews. This is common in affect corpus compilation. For instance, in the RECOLA [93] and SEWA [57] databases, experts annotate socio-affective data from participants involved in collaborative tasks or discussions.

Several studies exemplify these methods. In one, tension was annotated by two expert Hearthstone players using the PAGAN continuous annotation tool. The annotators assessed tension during competitive matches, considering gameplay features and players' facial reactions. This approach enabled highly accurate tension prediction, underscoring the value of expert annotations in understanding emotional responses in high-stakes gaming environments [65]. Similarly, interviews with highly engaged video game players in another study uncovered key motivational factors like socialization, challenge, and positive affect, crucial for understanding continued engagement [43].

However, these methods can suffer from variability in annotations due to differences in expertise and interpretation among annotators. In the Hearthstone study, despite the annotators' expertise, inconsistencies in perceived tension may have occurred due to their subjective understanding of the game's dynamics [65]. Additionally, [100] highlighted challenges in achieving consistency in participant responses during interviews. The open-ended nature of interviews can lead to varied interpretations of questions, and a language barrier with reaction cards in their engagement mapping method emphasized the need for careful design and testing of such tools [100].

4.1.3 Game-Design Enforced Engagement Level. Other approaches involve simulating various engagement levels through deliberate game-level design elements [75]. These controlled settings intentionally craft environments within games that range from highly stimulating to deliberately mundane or challenging, often used in adaptive game design [12]. This method manipulates player engagement by altering factors like visual complexity, difficulty, or narrative progression. However, a limitation of this approach is the assumption that players universally perceive each game-level design with the intended level of engagement. Individual player preferences, cognitive styles, and prior gaming experiences can significantly influence how these design elements are perceived, potentially leading to varied engagement levels that may not align perfectly with the designer's intentions. Thus, while valuable for exploring the impact of design choices on engagement, these methods must consider the diverse and subjective nature of player responses to game environments.

4.1.4 Proxy Methods. Proxy methods involve various approaches to understanding and quantifying player engagement indirectly. These methods use game design elements and observable behaviors as proxies for deeper psychological states, offering insights into engagement dynamics without relying solely on self-reports or observational methods. Their non-intrusive nature allows for broad applicability and easy implementation across diverse gaming contexts.

A prominent proxy method in the literature is *continuation desire*, or *conation*. This method measures a player's intrinsic motivation to persist in gameplay, seen as a fundamental indicator of engagement [87]. By assessing players' willingness to continue across different stages or levels, researchers quantify engagement through observable behaviors rather than explicit self-reports. Studies show that continuation desire strongly correlates with players' emotional states and interaction with game challenges [99]. Conation is linked to the *happy-gets-happier* effect, where initially engaged players become more engaged over time as they experience success and mastery within the game environment [89]. In games like Candy Crush Saga, engagement follows a power-law behavior, with players investing more time as they progress, illustrating the compounding effect of positive engagement [89]. This approach offers a quantitative measure of engagement and insights into how game progression and difficulty influence player motivation and commitment.

Table 4. Comparison of Player Engagement Validation Methods

Method	Papers	Real-Time	Objective	Comprehensive	Unbiased
QREs	[14, 23, 25, 27, 51, 68, 71, 77, 96, 103, 116]	✗	✓	✓	✗
Cont. Annot.	[57, 65, 70, 83, 93]	✓	✓	✓	✗
Interviews	[43, 100]	✗	✗	✓	✗
Level Design	[12, 75]	✗	✓	✗	✓
Proxy	[20, 87, 89, 99]	✓	✓	✗	✓

However, conation's narrow focus on goal-directed behavior and persistence may overlook other critical components of engagement, such as emotional immersion, cognitive absorption, or social interaction within the game. This limitation may not fully capture instances where players are emotionally invested but not necessarily driven by specific objectives, or where engagement fluctuates due to complex emotional responses during gameplay.

By analyzing how players interact during different gameplay phases—such as intense combat vs. downtime or exploration—researchers can infer engagement levels from behavioral cues and game state transitions. The FaceEngage annotation protocol [20] utilizes the picture-in-picture format of gameplay videos, categorizing game status (active or transitional) in the main screen and user play status (busy or idle) in the inset window. High engagement is inferred from active game status with busy user play, while low engagement is associated with transitional game status and idle user play. Uncertain cases are excluded to maintain data quality and reduce bias.

A limitation of the dataset's annotation method is that it may conflate player-audience engagement with player-game engagement, potentially skewing results. The protocol relies primarily on visible game states and player actions, without accounting for cognitive states or emotional responses that aren't immediately apparent. For instance, a player might appear engaged due to active gameplay but could be emotionally disconnected or distracted. This surface-level assessment may not fully represent the complex, multifaceted nature of player engagement, particularly in diverse gaming scenarios or when players interact with their audience.

4.2 Comparison and Lessons Learned

The comprehensive review of engagement ground truth methods reveals that each approach presents distinct tradeoffs: Questionnaires provide thorough assessment but interrupt gameplay, continuous annotations enable temporal tracking but suffer from recall bias, observational methods offer expert insights but introduce subjectivity, game-design approaches ensure controlled environments but may not generalize, and proxy methods enable unobtrusive measurement but can oversimplify engagement.

Our analysis through four key criteria (Table 4)—real-time capability, objectivity, comprehensiveness, and bias minimization—demonstrates that no single method provides a complete solution. Continuous annotation methods achieve the best balance, satisfying three criteria, while interviews face the most limitations. These findings yield crucial implications for engagement validation:

— Real-time measurement often conflicts with comprehensive assessment.
— Objective scales improve reliability but may oversimplify engagement complexity.
— Effective validation requires combining complementary methods.
— Validation protocols must actively address temporal aspects and bias mitigation.

These insights emphasize the need for multi-method approaches in both research methodology and practical game development, carefully balancing measurement accuracy with implementation feasibility.

5 Player Engagement Estimation Models

Now that we have reviewed the required concepts and their related literature about player engagement, including multimodal features and validation methods, we offer in this section a review of the state-of-the-art methods and models for the automatic estimation of player engagement. We focus on video games, highlighting trends in the literature. Approaches vary, employing combinations of signals and techniques such as hierarchical Bayesian models, **Convolutional Neural Networks (CNNs)** with **Long Short-Term Memory (LSTM)** networks, and other innovative methods. By examining these methodologies, we aim to elucidate their advantages, limitations, and unique contributions to understanding player engagement dynamics. We divide the existing estimation methods into three groups: face-based, physiological and behavioral-based, and multimodal techniques, described next.

5.1 Face-Based Models

These models aim to infer player engagement levels from facial expressions and, in some cases, incorporate audio data for improved accuracy. All recent models employ various machine learning techniques. Each approach presents unique strengths and limitations that highlight gaps in the current literature, discussed next.

5.1.1 FaceEngage Dataset. Chen et al. [20] introduced the FaceEngage dataset to address non-intrusive engagement estimation in gaming using user-contributed gameplay videos. They proposed two methods: (1) Feature Extraction and Traditional ML, which uses fixed-length feature vectors from facial motion data with traditional classifiers, and (2) Deep Learning Approach, employing a pre-trained VGG-Face CNN for face embedding extraction, followed by an encoder-decoder RNN with an attention mechanism for temporal processing. The deep learning approach significantly outperformed the traditional method, achieving an accuracy of 83.8%. This study demonstrates the potential of using facial expressions for non-intrusive engagement estimation in gaming contexts. However, it faces potential overfitting due to limited training data, especially for the deep learning model. Furthermore, it uses a simplistic annotation approach, explained in Section 4.1.4, which may not capture the full complexity of engagement, and the study does not consider the impact of individual differences in facial expressiveness or cultural variations in expression.

5.1.2 Facial Expressions for Conation Prediction. Rae et al. [87] developed a system to measure emotions and continuation desire (conation) during gameplay using two machine learning algorithms: an emotion recognition system trained on 2 million images, and a continuation desire predictor using LSTM networks. The integrated system processes facial expressions in real-time, making it suitable for live gaming environments. Validation on a different game showed 78.48% accuracy in predicting continuation desire, demonstrating some generalizability. However, the system's reliance on conation as a measure of engagement may not fully capture the nuances of emotional immersion or cognitive absorption in gameplay experiences. The study observed less pronounced emotional expressions in single-player games, which may limit the system's effectiveness in such contexts. Additionally, the approach does not account for engagement that may persist even when a player doesn't explicitly desire to continue playing.

These face-based engagement estimation models show promise in non-intrusive player engagement assessment, but they also highlight significant aspects in the current literature:

(1) *Non-Intrusive Measurement*: Both studies demonstrate the potential of using facial expressions for unobtrusive engagement estimation in gaming contexts.
(2) *Deep Learning Superiority*: The FaceEngage study shows that deep learning approaches significantly outperform traditional methods in this domain.
(3) *Real-Time Applicability*: The system by Rae et al. [87] processes facial expressions in real-time, making it suitable for live gaming environments.
(4) *Limited Engagement Definitions*: Both studies focus on narrow aspects of engagement without considering its multifaceted nature.
(5) *Contextual Challenges*: The models often struggle to account for varying gameplay contexts, such as single-player vs. multiplayer scenarios.
(6) *Ground Truth Reliability*: The simplistic annotation approaches may not capture the complexity of engagement.

Future research should address these limitations by developing more comprehensive engagement models that account for its multidimensional nature, consider diverse gaming contexts, and employ more sophisticated multimodal integration techniques. Additionally, establishing more reliable methods for ground truth annotation and exploring ways to personalize engagement estimation models could significantly advance the field.

5.2 Physiological and Behavioral-Based Models

Recent research has explored the use of physiological and behavioral data to estimate player engagement and enjoyment in gaming contexts. These approaches aim to provide more objective measures of engagement compared to traditional self-report methods, potentially offering real-time insights into player experiences.

5.2.1 Predicting Fun from Physiological and Behavioral Data. Fortin et al. [35] developed real-time engagement models using diverse data sources and machine learning techniques, aiming to improve adaptive gaming systems by accurately assessing player enjoyment. The study collected comprehensive data from 218 participants playing Ubisoft's Assassin's Creed games. The researchers gathered an extensive array of data, including physiological measures (ECG, respiratory activity, EDA, EMG, eye movements), questionnaires, game events, and continuous fun ratings. They extracted 244 features from this data, divided into time-dependent and time-independent categories. Initially, regression models struggled to predict fun ratings due to noisy labels. The researchers then shifted to classification methods, categorizing fun into low, neutral, and high states. Among various classifiers tested, the XGBoost classifier outperformed others, achieving the highest F1 score. This classifier highlighted the importance of features from respiratory activity, ECG, eye tracking, questionnaires, and EMG. The study found that the XGBoost classifier was particularly effective at predicting high fun states but less so for neutral fun states. This may be due to participants generally reporting high levels of fun during gameplay. An alternative ranking method, which classified changes in fun levels instead of predicting absolute values, did not improve accuracy and increased label noise. Despite its comprehensive approach, the study's high proportion of positive fun ratings may have biased results, and the imbalanced gender distribution (184 males vs. 9 females) could affect feature importance and generalizability.

5.2.2 Bayesian Hierarchical Models for Motivation Prediction. Sawyer et al. [97] explored the use of Bayesian hierarchical models to predict player motivation from in-game actions in educational interactive narrative games. Using the Crystal Island game, they modeled engagement across multiple contexts, capturing both general trends and specific differences between player groups.

The Bayesian hierarchical linear model, trained using Markov Chain Monte Carlo sampling, outperformed pooled and context-specific models in predicting player motivation. This approach proved particularly effective in diverse classroom settings where context significantly affects engagement. The model's ability to provide posterior distributions offered valuable insights into how various game features influence engagement across different demographics and environments. However, data variability and availability limited the model's generalizability to broader populations or new game versions.

These above two physiological and behavioral data-based engagement estimation models demonstrate promising approaches to understanding player engagement, highlighting several key points and challenges:

- *Multimodal Data Integration*: The study by Fortin et al. [35] showcases the potential of combining various physiological and behavioral signals for engagement estimation.
- *Context-Aware Modeling*: The Bayesian hierarchical model by Sawyer et al. [97] demonstrates the importance of considering different contexts and player groups in engagement modeling.
- *Temporal Dynamics*: Both studies highlight the challenge of capturing rapid fluctuations in engagement during gameplay.
- *Demographic Considerations*: The studies emphasize the need for more diverse and representative participant pools in future research.
- *Generalizability*: Both studies were limited to specific games or genres, raising questions about model applicability across different types of games.
- *Balancing Objectivity and Subjectivity*: While physiological measures offer objective data, correlating them with subjective experiences of engagement remains challenging.

Future research in this area should focus on developing more robust, generalizable models that can account for individual differences and diverse gaming contexts. Additionally, exploring ways to combine physiological and behavioral data with other engagement indicators, such as facial expressions or voice analysis, could provide a more comprehensive understanding of player engagement.

5.2.3 Industry Approaches. While academic research focuses on sophisticated modeling techniques, commercial platforms offer streamlined approaches to engagement estimation. For example, GameAnalytics' event-based tracking system allows developers to define custom triggers for specific player actions like completing challenging levels or making in-game purchases, enabling precise correlation with long-term engagement [45]. Their platform tracks 22 key metrics across engagement, monetization, and advertising categories, including granular measurements of session length, player count, and churn rate [44]. Similarly, SonaMine's analytics suite emphasizes metrics that directly tie to business outcomes, such as the stickiness ratio (DAU/MAU) for measuring regular player engagement, and detailed conversion tracking for monetization effectiveness [58]. These industry tools prioritize actionable insights—for instance, allowing developers to identify exactly which game levels or features correlate with higher retention rates [1]. This data-driven approach complements academic research by providing specific, measurable validations of engagement theories and highlighting concrete areas where theoretical models can improve game design and player retention.

5.3 Multimodal Techniques

Recent research has explored multimodal approaches beyond the traditional single-mode methods of incorporating physiological signals. In the following two such multimodal approaches are described.

5.3.1 Face, Pixel, and Audio. Pan et al. [79] extended the FaceEngage dataset by proposing a multimodal deep learning model that incorporates facial, pixel, and sound modalities. Their approach includes a face modality using a Multi-Task Cascaded CNN and EfficientNet-B0 [105], a pixel modality processing entire video frames, and a sound modality extracting features from Mel-Frequency Cepstral coefficients. The model achieves 77.2% accuracy, demonstrating the potential of multimodal approaches. Notably, the study quantifies modality contributions, revealing that the sound modality dominates with a 92.6% contribution. However, this overwhelming contribution raises questions about the necessity and efficiency of including face and pixel modalities. Furthermore, the dataset's annotation method may conflate player-audience engagement with player-game engagement, potentially skewing results, as explained in Section 4.1.4. The study also does not explore more complex fusion strategies, which might yield better performance.

5.3.2 Gamepad and Pixels. Pinitas et al. [83] introduced a multimodal approach to predict player engagement in Tom Clancy's The Division 2, using gameplay frames and gamepad inputs. Their novel dataset comprised annotated gameplay videos and gamepad actions from 25 participants. The methodology involved processing gameplay frames with a pre-trained ResNet18 neural network and encoding gamepad actions into frequencies and combos. Separate neural network architectures were used for each modality, with a fusion model integrating features from both sources. Evaluation using leave-2-participants-out cross-validation showed that the fusion model, combining both modalities, exhibited the highest accuracy in predicting player engagement. This validated the effectiveness of multimodal approaches for engagement prediction. A notable limitation was recall bias from the gap between gameplay engagement and annotation, potentially affecting the fidelity of engagement traces.

These multimodal techniques for engagement estimation in gaming highlight several key areas for future research:

- *Modality Contributions*: The study by Pan et al. [79] quantifies the relative importance of different modalities, revealing the dominance of audio in their context. These findings raise questions about the necessity of including less contributive modalities in engagement models.
- *Multimodal Integration*: Both studies showcase the potential of combining multiple data sources for more accurate engagement prediction, but also highlight the challenges in effective integration.
- *Generalizability Challenges*: Each study faces limitations in terms of sample size, game specificity, or data collection methods, indicating a common challenge in the field.
- *Recall Bias Challenge*: The study by Pinitas et al. exemplifies the challenge of recall bias in establishing reliable engagement annotations, a common issue in engagement research.

5.4 Comparison and Lessons Learned

The review of engagement estimation models (Table 5) reveals distinct patterns across three modeling approaches. Face-based models achieve high accuracy (78–84%) but often rely on simplified engagement definitions. Physiological and behavioral models offer objective measurements but struggle with data imbalance and limited generalizability. Multimodal approaches show promise in combining complementary signals but face challenges in effective feature integration and annotation quality.

This systematic comparison yields several critical insights for engagement modeling:

- Model complexity often trades off with interpretability—deep learning approaches outperform traditional methods but offer less insight into feature importance.

Table 5. Summary of Player Engagement Estimation Models in Video Games

Type	Study	Features	Target	Data	Model	Performance	Limitation
Face-Based	Chen et al. [20]	Face embed., motion	Binary Eng.	700+ videos	VGG-Face+IRNN w/attention	83.8% Acc.	Simplistic annotation
	Selvig and Schoenau-Fog [87]	Face embed., emotions	Conation	2.6M points	LSTM	78.48% Acc.	Limited definition
Physio. and Behav.	Fortin-Côté et al. [35]	Physiological	Three-level fun	218 part.	XGBoost	0.38 F1	Data imbalance
	Sawyer et al. [97]	Gameplay Features	Motivation	63 students	Bayesian Hierarchical	1.469 MSE	Limited generalizability
Multi-Modal	Pan et al. [79]	Face, pixel, audio	Binary Eng.	700+ videos	EfficientNet + GRU, Audio CNN	77.2% Acc.	Simplistic annotation
	Pinitas et al. [83]	Game pixels, user inputs	Binary Eng.	25 part.	ResNet18 + NN	72% Acc.	Recall bias

— Ground truth quality significantly impacts model performance, with simplified annotations and recall bias limiting accuracy.
— Context-awareness remains challenging, with most models showing limited generalizability across different games or player populations.
— Multimodal integration requires careful consideration of modality contributions, as demonstrated by the dominance of audio features in some studies.

These findings suggest that future engagement modeling should prioritize robust ground truth collection, context-aware architectures, and thoughtful multimodal integration while maintaining interpretability.

6 Conclusions and Future Work

The subjective, multi-dimensional nature of player engagement has led to various research strands, each addressing player engagement measurement from different perspectives. In this article, we reviewed concepts, predictors, validation methods, and estimation models of player engagement, aiming to highlight common challenges and uncover opportunities for more coherent, mature research directions. In this section, we briefly cover the conclusion points and future work directions. For expanded summary, conclusions, and future works, please see Appendix B.

Our review has identified four crucial dimensions for advancing player engagement estimation in video games. First, the conceptual framework must recognize engagement as a dynamic process spanning cognitive, affective, and behavioral dimensions, where constructs progress from motivation through immersion and flow to endurability. Second, while each measurement modality offers unique perspectives—physiological signals provide objective arousal measures, EEG captures neural responses, facial expressions reveal emotions, eye tracking shows attention patterns, and gameplay data reflects behavior—multimodal approaches yield the most comprehensive understanding by compensating for individual limitations. Third, ground truth establishment remains challenging due to engagement's subjective nature, with an optimal validation method requiring continuous, non-disruptive self-reporting that balances objectivity with player experience preservation. Finally, current estimation models, though promising, reveal critical tradeoffs between complexity and interpretability, accuracy and generalizability, with industry practices emphasizing practical metrics like DAU/MAU ratios, session length, and custom engagement triggers, while academic research explores more sophisticated but less scalable techniques. Future research should prioritize developing context-aware, computationally efficient algorithms that process heterogeneous data streams in real-time, with academia-industry collaboration facilitating access to diverse datasets, practical implementation contexts, and real-world validation environments to bridge the gap between sophisticated research models and actionable, scalable engagement estimation methods.

References

[1] Player Engagement Metrics—larksuite.com. 2024. Retrieved February 22, 2025 from https://www.larksuite.com/en_us/topics/gaming-glossary/player-engagement-metrics
[2] Video Game Market Report. 2024. Retrieved February 22, 2025 from https://www.globalgrowthinsights.com/market-reports/video-game-market-107798
[3] Amir Zaib Abbasi, Ding Hooi Ting, and Helmut Hlavacs. 2017. Engagement in games: Developing an instrument to measure consumer videogame engagement and its validation. *International Journal of Computer Games Technology* 2017, 1 (2017), 7363925.
[4] Krishan Arora. 2023. The Gaming Industry: A Behemoth with Unprecedented Global Reach.
[5] Christopher Bailey, Elaine Pearson, Stavroula Gkatzidou, and Steve Green. 2006. Using video games to develop social, collaborative and communication skills. In *EdMedia+ Innovate Learning*. Association for the Advancement of Computing in Education (AACE), 1154–1161.
[6] Marta Bassi and Antonella Delle Fave. 2016. *Flow in the Context of Daily Experience Fluctuation*. Springer International Publishing, Cham, 181–196.
[7] Michael Bennett, Lukas Čironis, Amber Sousa, Sophia L. Ahmad, Tamzid Hassan, Kyle Yuen, Peter Douris, Hallie Zwibel, Joanne DiFrancisco-Donoghue, and Aaron Koshy. 2022. Continuous monitoring of HRV in esports players. *International Journal of Esports* 1, 1 (2022), 14 pages.
[8] Mehmet Ilker Berkman and Ecehan Akan. 2019. *Presence and Immersion in Virtual Reality*. Springer International Publishing, Cham, 1–10.
[9] Riccardo Berta, Francesco Bellotti, Alessandro De Gloria, Danu Pranantha, and Carlotta Schatten. 2013. Electroencephalogram and physiological signal analysis for assessing flow in games. *IEEE Transactions on Computational Intelligence and AI in Games* 5, 2 (2013), 164–175.
[10] N. Bianchi-Berthouze. 2013. Understanding the role of body movement in player engagement. *Human–Computer Interaction* 28, 1 (2013), 40–75.
[11] Thomas Bjørner. 2023. Using EEG data as dynamic difficulty adjustment in a serious game about the plastic pollution in the oceans. In *2023 ACM Conference on Information Technology for Social Good (GoodIT '23)*. ACM, New York, NY, 6–15.
[12] Boyan Bontchev and Dessislava Vassileva. 2016. Assessing engagement in an emotionally-adaptive applied game. In *4th International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM '16)*. ACM, New York, NY, 747–754.
[13] Cheryl Campanella Bracken and Paul Skalski. 2006. Presence and video games: The impact of image quality and skill level. In *9th Annual International Workshop on Presence*. Cleveland State University, OH, 28–29.
[14] Jeanne H. Brockmyer, Christine M. Fox, Kathleen A. Curtiss, Evan McBroom, Kimberly M. Burkhart, and Jacquelyn N. Pidruzny. 2009. The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology* 45, 4 (2009), 624–634.
[15] Emily Brown and Paul Cairns. 2004. A grounded investigation of game immersion. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems (CHI EA '04)*. ACM, New York, NY, 1297–1300.
[16] Michael Burch and Kuno Kurzhals. 2020. Visual analysis of eye movements during game play. In *ACM Symposium on Eye Tracking Research and Applications (ETRA '20 Short Papers)*. ACM, New York, NY, Article 59, 5 pages.
[17] Loïc Caroux. 2023. Presence in video games: A systematic review and meta-analysis of the effects of game design choices. *Applied Ergonomics* 107 (2023), 103936.
[18] Guillaume Chanel, Cyril Rebetez, Mireille Bétrancourt, and Thierry Pun. 2008. Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. In *12th International Conference on Entertainment and Media in the Ubiquitous Era (MindTrek '08)*. ACM, New York, NY, 13–17.
[19] Jenova Chen. 2007. Flow in games (and everything else). *Communications of the ACM* 50, 4 (Apr. 2007), 31–34.
[20] Xu Chen, Li Niu, Ashok Veeraraghavan, and Ashutosh Sabharwal. 2019. FaceEngage: Robust estimation of gameplay engagement from user-contributed (YouTube) videos. *IEEE Transactions on Affective Computing* 13, 2 (2019), 651–665.
[21] Anders Christofferson, Anders Videbaek, Alex Egan, Tom Rowland, and Matt Madden. 2024. Gamer Survey: Young Players Reshape the Industry. Retrieved February 22, 2025 from https://www.bain.com/insights/gamer-survey-young-players-reshape-the-industry-gaming-report-2024/

[22] Ben Cowley, Darryl Charles, Michaela Black, and Ray Hickey. 2008. Toward an understanding of flow in video games. *Computers in Entertainment (CIE)* 6, 2 (2008), 1–27.

[23] Mihaly Csikszentmihalyi. 1988. The flow experience and its significance for human psychology. *Optimal Experience: Psychological Studies of Flow in Consciousness* 2 (1988), 15–35.

[24] Mihaly Csikszentmihalyi. 1990. *Flow: The Psychology of Optimal Experience*. Harper Perennial, New York, NY.

[25] Yvonne A. W. De Kort, Wijnand A. IJsselsteijn, and Karolien Poels. 2007. Digital games as social presence technology: Development of the social presence in gaming questionnaire (SPGQ). *Proceedings of PRESENCE* 195203 (2007), 1–9.

[26] W. N. Dember and L. Penwell. 1980. Happiness, depression, and the Pollyanna principle. *Bulletin of the Psychonomic Society* 15, 5 (1980), 321–323.

[27] Alena Denisova, A. Imran Nordin, and Paul Cairns. 2016. The convergence of player experience questionnaires. In *2016 Annual Symposium on Computer-Human Interaction in Play*. ACM, New York, NY, 33–37.

[28] K. Doherty and G. Doherty. 2018. Engagement in HCI: Conception, theory and measurement. *ACM Computing Surveys* 51, 5 (2018), 1–39.

[29] Anders Drachen, Alessandro Canossa, and Georgios N. Yannakakis. 2009. Player modeling using self-organization in Tomb Raider: Underworld. In *2009 IEEE Symposium on Computational Intelligence and Games*. IEEE, Milan, Italy, 1–8.

[30] Haihan Duan, Yiwei Huang, Yifan Zhao, Zhen Huang, and Wei Cai. 2022. User-generated content and editors in video games: Survey and vision. In *2022 IEEE Conference on Games (CoG)*. IEEE, 536–543.

[31] Emotiv. 2023. Performance Metrics. Retrieved February 22, 2025 from https://www.emotiv.com/pages/performance-metrics

[32] Laura Ermi and Frans Mäyrä. 2007. Fundamental components of the gameplay experience: Analyzing immersion. *Worlds in Play: International Perspectives on Digital Games Research* 21 (2007), 37.

[33] Stephen Fairclough. 2007. Psychophysiological inference and physiological computer games. In *Brain-Computer Interfaces and Games Workshop at Advances in Computer Entertainment (AcE)*, 6.

[34] Lisa A. Feldman. 1995. Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of Personality and Social Psychology* 69, 1 (1995), 153.

[35] Alexis Fortin-Côté, Cindy Chamberland, Mark Parent, Sébastien Tremblay, Philip Jackson, Nicolas Beaudoin-Gagnon, Alexandre Campeau-Lecours, Jérémy Bergeron-Boucher, and Ludovic Lefebvre. 2019. Predicting video game players' fun from physiological and behavioural data. In *Advances in Information and Communication Networks*. Springer International Publishing, Cham, 479–495.

[36] Daniel Gábana Arellano, Laurissa Tokarchuk, and Hatice Gunes. 2017. Measuring affective, physiological and behavioural differences in solo, competitive and collaborative games. In *Intelligent Technologies for Interactive Entertainment*. Springer International Publishing, Cham, 184–193.

[37] Alan Gevins, Michael E. Smith, Linda McEvoy, and Daphne Yu. 1997. High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex (New York, NY: 1991)* 7, 4 (1997), 374–385.

[38] Marco Granato, Davide Gadia, Dario Maggiorini, and Laura Anna Ripamonti. 2017. Emotions detection through the analysis of physiological information during video games fruition. In *Games and Learning Alliance*. Springer International Publishing, Cham, 197–207.

[39] Gianluca Guglielmo, Paris Mavromoustakos Blom, Michal Klincewicz, Boris Čule, and Pieter Spronck. 2022. Face in the game: Using facial action units to track expertise in competitive video game play. In *2022 IEEE Conference on Games (CoG)*. IEEE, Beijing, China, 112–118.

[40] Tehmina Hafeez, Sanay Muhammad Umar Saeed, Aamir Arsalan, Syed Muhammad Anwar, Muhammad Usman Ashraf, and Khalid Alsubhi. 2021. EEG in game user analysis: A framework for expertise classification during gameplay. *PLoS One* 16, 6 (2021), e0246913.

[41] Eman Hassan. 2006. Recall bias can be a threat to retrospective and prospective research designs. *International Journal of Epidemiology* 3, 2 (2006), 339–412.

[42] Richard L. Hazlett. 2006. Measuring emotional valence during interactive experiences: Boys at video game play. In *SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, 1023–1026.

[43] Bobby Hoffman and Louis Nadelson. 2010. Motivational engagement and video gaming: A mixed methods study. *Educational Technology Research and Development* 58 (2010), 245–270.

[44] Tomas Hubka. 2024. 22 Metrics All Game Developers Should Know by Heart—gameanalytics.com. Retrieved February 22, 2025 from https://gameanalytics.com/blog/metrics-all-game-developers-should-know/

[45] Tomas Hubka. 2024. Decoding Players' Patterns with Engagement Tracing—gameanalytics.com. Retrieved February 22, 2025 from https://gameanalytics.com/blog/engagement-tracing-retention

[46] Sinh Huynh, Seungmin Kim, JeongGil Ko, Rajesh Krishna Balan, and Youngki Lee. 2018. Engagemon: Multi-modal engagement sensing for mobile games. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–27.

[47] W. A. IJsselsteijn, Y. A. W. de Kort, and K. Poels. 2013. *The Game Experience Questionnaire*. Technische Universiteit Eindhoven, Eindhoven, NL.

[48] Institute for Operations Research and the Management Sciences. 2019. New research analyzes video game player engagement. *ScienceDaily* (25 September 2019). Retrieved from www.sciencedaily.com/releases/2019/09/190925133637.htm

[49] Faisal Iqbal. 2015. Understanding User Interaction in a Video Game by Using Eye Tracking and Facial Expressions Analysis. Master's thesis. University of Tampere, School of Information Sciences.

[50] Katherine Isbister and Noah Schaffer. 2008. *Game Usability: Advancing the Player Experience*. Morgan Kaufmann, Boston.

[51] Charlene Jennett, Anna L. Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies* 66, 9 (2008), 641–661.

[52] Joshua Juvrud, Gabriel Ansgariusson, Patrik Selleby, and Magnus Johansson. 2021. Game or watch: The effect of interactivity on arousal and engagement in video game media. *IEEE Transactions on Games* 14, 2 (2021), 308–317.

[53] A. T. Kamzanova, Gerald Matthews, A. M. Kustubayeva, and S. M. Jakupov. 2011. EEG indices to time-on-task effects and to a workload manipulation (cueing). *International Journal of Psychological and Behavioral Sciences* 5, 8 (2011), 928–931.

[54] Tanaya Killedar, Pragya, Gunjan Surya, and Maheshwari Rathod. 2021. Fuzzy logic for video game engagement analysis using facial emotion recognition. In *2021 8th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, 481–485.

[55] Jung-Yong Kim, Hea-Sol Kim, Dong-Joon Kim, Sung-Kyun Im, and Mi-Sook Kim. 2021. Identification of video game addiction using heart-rate variability parameters. *Sensors* 21, 14 (2021), 4683.

[56] Isabelle Kniestedt, Iulia Lefter, Stephan Lukosch, and Frances M. Brazier. 2022. Re-framing engagement for applied games: A conceptual framework. *Entertainment Computing* 41 (2022), 100475.

[57] Jean Kossaifi, Robert Walecki, Yannis Panagakis, Jie Shen, Maximilian Schmitt, Fabien Ringeval, Jing Han, Vedhas Pandit, Antoine Toisoul, Björn Schuller, et al. 2019. Sewa db: A rich database for audio-visual emotion and sentiment research in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 3 (2019), 1022–1040.

[58] Nick Lim. 2024. 10 Types of Game Metrics and How to Use Them—sonamine.com. Retrieved February 22, 2025 from https://www.sonamine.com/blog/10-types-of-game-metrics-and-how-to-use-them

[59] Phil Lopes, Georgios N. Yannakakis, and Antonios Liapis. 2017. RankTrace: Relative and unbounded affect annotation. In *2017 7th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, TX, 158–163.

[60] Wenyi Lu, Hao He, Alex Urban, and Joe Griffin. 2021. What the eyes can tell: Analyzing visual attention with an educational video game. In *ACM Symposium on Eye Tracking Research and Applications (ETRA '21 Short Papers)*. ACM, New York, NY, Article 36, 7 pages.

[61] Konstantinos Makantasis, Antonios Liapis, and Georgios N. Yannakakis. 2019. From pixels to affect: A study on games and player experience. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, Cambridge, UK, 1–7.

[62] Konstantinos Makantasis, Antonios Liapis, and Georgios N. Yannakakis. 2021. The pixels and sounds of emotion: General-purpose representations of arousal in games. *IEEE Transactions on Affective Computing* 14, 1 (2021), 680–693.

[63] Regan L. Mandryk and Kori M. Inkpen. 2004. Physiological indicators for the evaluation of co-located collaborative play. In *2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04)*. ACM, New York, NY, 102–111.

[64] Paris Mavromoustakos-Blom, Mehmet Kosa, Sander Bakkes, and Pieter Spronck. 2021. Correlating facial expressions and subjective player experiences in competitive hearthstone. In *16th International Conference on the Foundations of Digital Games (FDG '21)*. ACM, New York, NY, Article 41, 5 pages.

[65] Paris Mavromoustakos-Blom, Dávid Melhárt, Antonios Liapis, Georgios N. Yannakakis, Sander Bakkes, and Pieter Spronck. 2023. Multiplayer tension in the wild: A hearthstone case. In *18th International Conference on the Foundations of Digital Games (FDG '23)*. ACM, New York, NY, Article 16, 9 pages.

[66] Alison McMahan. 2013. Immersion, engagement, and presence: A method for analyzing 3-D video games. In *The Video Game Theory Reader*, 67–86.

[67] Albert Mehrabian. 1980. *Basic Dimensions for a General Psychological Theory: Implications for Personality, Social, Environmental, and Developmental Studies*. Oelgeschlager, Gunn & Hain.

[68] David Melhart, Ahmad Azadvar, Alessandro Canossa, Antonios Liapis, and Georgios N. Yannakakis. 2019. Your gameplay says it all: Modelling motivation in Tom Clancy's the division. In *2019 IEEE Conference on Games (CoG)*. IEEE, London, UK, 1–8.

[69] David Melhart, Daniele Gravina, and Georgios N. Yannakakis. 2020. Moment-to-moment engagement prediction through the eyes of the observer: PUBG streaming on Twitch. In *15th International Conference on the Foundations of Digital Games (FDG '20)*. ACM, New York, NY, Article 60, 10 pages.

[70] David Melhart, Antonios Liapis, and Georgios N. Yannakakis. 2022. The arousal video game annotation (AGAIN) dataset. *IEEE Transactions on Affective Computing* 13, 4 (2022), 2171–2184.

[71] Antonio Miguel-Cruz, Adriana Maria Rios Rincon, Christine Daum, Daniel Alejandro Quiroga Torres, Ruby De Jesus, Lili Liu, and Eleni Stroulia. 2021. Predicting engagement in older adults with and without dementia while playing mobile games. *IEEE Instrumentation & Measurement Magazine* 24, 6 (2021), 29–36.

[72] Josh Aaron Miller, Uttkarsh Narayan, Matthew Hantsbarger, Seth Cooper, and Magy Seif El-Nasr. 2019. Expertise and engagement: Re-designing citizen science games with players' minds in mind. In *14th International Conference on the Foundations of Digital Games (FDG '19)*. ACM, New York, NY, Article 6, 11 pages.

[73] Mahtab Mohammadpoor Faskhodi, Mireya Fernández-Chimeno, and Miquel Angel García-González. 2023. Arousal detection by using ultra-short-term heart rate variability (HRV) analysis. *Frontiers in Medical Engineering* 1 (2023), 1209252.

[74] Giovanni B. Moneta. 2012. *On the Measurement and Conceptualization of Flow*. Springer, New York, NY, 23–50.

[75] Lennart Nacke and Craig Lindley. 2009. Affective ludology, flow and immersion in a first-person shooter: Measurement of player experience. *Loading..., The Journal of the Canadian Game Studies Association* 3, 5 (2009), 21.

[76] Manuel Ninaus, Kristian Kiili, Guilherme Wood, Korbinian Moeller, and Silvia Erika Kober. 2020. To add or not to add game elements? Exploring the effects of different cognitive task designs using eye tracking. *IEEE Transactions on Learning Technologies* 13, 4 (2020), 847–860.

[77] A. I. Nordin, A. Denisova, and P. Cairns. 2014. Too many questionnaires: Measuring player experience whilst playing digital games. In *the 7th York Doctoral Symposium on Computer Science and Electronics*. University of York, York, UK, 6.

[78] Heather L. O'Brien and Elaine G. Toms. 2008. What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology* 59, 6 (2008), 938–955.

[79] Sicheng Pan, Gary J. W. Xu, Kun Guo, Seop Hyeong Park, and Hongliang Ding. 2023. Video-based engagement estimation of game streamers: An interpretable multimodal neural network approach. *IEEE Transactions on Games* (2023), 1–12.

[80] Thomas D. Parsons, Timothy McMahan, and Ian Parberry. 2020. Classification of video game player experience using consumer-grade electroencephalography. *IEEE Transactions on Affective Computing* 13, 1 (2020), 3–15.

[81] Sarah Parvini. 2024. Insights into the artificial intelligence AI in video games market's growth potential 2024-2033. *AP News* (2024). Retrieved from https://apnews.com/article/c1327bb9130136d0a5f658f44176c5e7

[82] Jorge Peña and Jeffrey T. Hancock. 2006. An analysis of socioemotional and task communication in online multiplayer video games. *Communication Research* 33, 1 (2006), 92–109.

[83] Kosmas Pinitas, David Renaudie, Mike Thomsen, Matthew Barthet, Konstantinos Makantasis, Antonios Liapis, and Georgios N. Yannakakis. 2023. Predicting player engagement in Tom Clancy's The Division 2: A multimodal approach via pixels and gamepad actions. In *25th International Conference on Multimodal Interaction (ICMI '23)*. ACM, New York, NY, 488–497.

[84] C. Politowski, Fabio Petrillo, and Yann-Gaël Guéhéneuc. 2020. Improving engagement assessment in gameplay testing sessions using IoT sensors. In *IEEE/ACM 42nd International Conference on Software Engineering Workshops (ICSEW '20)*. ACM, New York, NY, 655–659.

[85] Emma J. Pretty, Renan Guarese, Chloe A. Dziego, Haytham M. Fayek, and Fabio Zambetta. 2024. Multimodal measurement of cognitive load in a video game context: A comparative study between subjective and objective metrics. *IEEE Transactions on Games* (2024), 1–14.

[86] K. Procci. 2015. The Subjective Gameplay Experience: An Examination of the Revised Game Engagement Model. Ph.D. Dissertation. University of C.F.

[87] Dines Rae Selvig and Henrik Schoenau-Fog. 2020. Non-intrusive measurement of player engagement and emotions—Real-time deep neural network analysis of facial expressions during game play. In *HCI in Games*. Springer International Publishing, Cham, 330–349.

[88] Janet C. Read, Sarah MacFarlane, and Chris Casey. 2002. Endurability, engagement and expectations: Measuring children's fun. *Interaction Design and Children* 2 (Aug. 2002), 1–23.

[89] David Reguera, Pol Colomer-de Simón, Iván Encinas, Manel Sort, Jan Wedekind, and Marián Boguñá. 2020. Quantifying human engagement into playful activities. *Scientific Reports* 10, 1 (2020), 4145.

[90] Simão Reis, Luís Paulo Reis, and Nuno Lau. 2019. Player engagement enhancement with video games. In *New Knowledge in Information Systems and Technologies*. Springer International Publishing, Cham, 263–272.

[91] Izabela Rejer and Michal Twardochleb. 2018. Gamers' involvement detection from EEG data with cGAAM–A method for feature selection for clustering. *Expert Systems with Applications* 101 (2018), 196–204.

[92] Ryan K. Rigney. 2024. The evolution of sponsored streaming in the gaming industry. *Polygon*. Retrieved from https://www.polygon.com/analysis/467688/push-to-talk-twitch-sponsored-stream-evolution

[93] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne. 2013. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE, Shanghai, China, 1–8.

[94] Ghulam Ruqeyya, Tehmina Hafeez, Sanay Muhammad Umar Saeed, and Aleeza Ishwal. 2022. EEG-based engagement index for video game players. In *International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE)*. IEEE, Lahore, Pakistan, 1–6.

[95] Richard M. Ryan and Edward L. Deci. 2000. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology* 25, 1 (2000), 54–67.

[96] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. 2006. The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion* 30 (2006), 344–360.

[97] Robert Sawyer, Jonathan Rowe, Roger Azevedo, and James Lester. 2018. Modeling player engagement with Bayesian hierarchical models. In *AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 257–263.

[98] Gianluca Schiavo, Alessandro Cappelletti, and Massimo Zancanaro. 2014. Engagement recognition using easily detectable behavioral cues. *Intelligenza Artificiale* 8, 2 (2014), 197–210.

[99] Henrik Schoenau-Fog. 2011. The player engagement process–An exploration of continuation desire in digital games. In *DiGRA 2011 Conference: Think Design Play*. Digital Games Research Association (DiGRA), Hilversum, The NL, 18.

[100] Henrik Schoenau-Fog and Thomas Bjørner. 2012. "Sure, I would like to continue": A method for mapping the experience of engagement in video games. *Bulletin of Science, Technology & Society* 32, 5 (2012), 405–412.

[101] Donghee Shin. 2019. How does immersion work in augmented reality games? A user-centric view of immersion and engagement. *Information, Communication & Society* 22, 9 (2019), 1212–1229.

[102] Katja Strelcova and Simay Karaağaç. 2024. Gaming: Multi-platform experiences and AI integration. *Think with Google* (2024). Retrieved from https://www.thinkwithgoogle.com/intl/en-emea/marketing-strategies/app-and-mobile/gaming-multi-platform-ai/

[103] Penelope Sweetser and Peta Wyeth. 2005. GameFlow: A model for evaluating player enjoyment in games. *Computers in Entertainment* 3, 3 (7 2005), 3.

[104] Mariusz Szwoch and Wioleta Szwoch. 2015. Emotion recognition for affect aware video games. In *Image Processing & Communications Challenges 6*. Springer International Publishing, Cham, 227–236.

[105] Mingxing Tan and Quoc Le. 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. In *36th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 97)*. PMLR, CA, 6105–6114.

[106] S. Tekofsky, P. H. M. Spronck, A. Plaat, H. J. van den Herik, and J. Broersen. 2013. PsyOps: Personality assessment through gaming behavior. In *the BNAIC Conference 2013*. Technische Universiteit, NL, 354–355.

[107] Simone Tognetti, Maurizio Garbarino, Andrea Bonarini, and Matteo Matteucci. 2010. Modeling enjoyment preference from physiological responses in a car racing game. In *2010 IEEE Conference on Computational Intelligence and Games*. IEEE, Copenhagen, Denmark, 321–328.

[108] Antoine Toisoul, Jean Kossaifi, Adrian Bulat, Georgios Tzimiropoulos, and Maja Pantic. 2021. Estimation of continuous valence and arousal levels from faces in naturalistic conditions. *Nature Machine Intelligence* 3, 1 (2021), 42–50.

[109] Shuo Wang, Xiaocao Xiong, Yan Xu, Chao Wang, Weiwei Zhang, Xiaofeng Dai, and Dongmei Zhang. 2006. Face-tracking as an augmented input in video games: enhancing presence, role-playing and control. In *SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, 1097–1106.

[110] Dustin T. Weiler, Stefanie O. Villajuan, Laura Edkins, Sean D. Cleary, and Jason J. Saleem. 2017. Wearable heart rate monitor technology accuracy in research: A comparative study between PPG and ECG technology. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61 (Sep. 2017), 1292–1296.

[111] Matthew R. Welsh, Emma Mosley, Sylvain Laborde, Melissa C. Day, Benjamin T. Sharpe, Rachel A. Burkill, and Phil D. J. Birch. 2023. The use of heart rate variability in esports: A systematic review. *Psychology of Sport and Exercise* 69 (2023), 102495.

[112] Eric N. Wiebe, Allison Lamb, Megan Hardy, and David Sharek. 2014. Measuring engagement in video game-based environments: Investigation of the user engagement scale. *Computers in Human Behavior* 32 (2014), 123–132.

[113] Joseph Wiggins, Mayank Kulkarni, Wookhee Min, Bradford Mott, Kristy Boyer, Eric Wiebe, and James Lester. 2018. Affect-based early prediction of player mental demand and engagement for educational games. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* 14, 1 (Sep. 2018), 243–249.

[114] Daniel Wilcox-Netepczuk. 2013. Immersion and realism in video games—The confused moniker of video game engrossment. In *CGAMES '2013 USA*. IEEE, KY, 92–95.

[115] A. Winklbauer, B. Stiglbauer, M. Lankes, and M. Sporn. 2023. Telling eyes: Linking eye-tracking indicators to affective variables. In *18th International Conference on the Foundations of Digital Games (FDG '23)*. ACM, New York, NY, Article 19, 11 pages.

[116] Bob G. Witmer and Michael J. Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence* 7, 3 (1998), 225–240.

[117] Erroll Wood, Tadas Baltruaitis, Xucong Zhang, Yusuke Sugano, Peter Robinson, and Andreas Bulling. 2015. Rendering of eyes for eye-shape registration and gaze estimation. In *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE, Santiago, Chile, 3756–3764.

[118] Wenlu Yang, Maria Rifqi, Christophe Marsala, and Andrea Pinna. 2018. Towards better understanding of player's game experience. In *2018 ACM on International Conference on Multimedia Retrieval (ICMR '18)*. ACM, New York, NY, 442–449.

[119] G. N. Yannakakis and J. Togelius. 2011. Experience-driven procedural content generation. *IEEE Transactions on Affective Computing* 2, 3 (2011), 147–161.

[120] Georgios N. Yannakakis, Pieter Spronck, Daniele Loiacono, and Elisabeth André. 2013. Player modeling. In *Artificial and Computational Intelligence in Games. Dagstuhl Follow-Ups*, Vol. 6. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 45–59.

[121] Judith Lynne Zaichkowsky. 1985. Measuring the involvement construct. *Journal of Consumer Research* 12, 3 (1985), 341–352.

Appendices

A A Player Engagement Questionnaire

Table A1. Questionnaires Measuring Player Engagement in Video Games [77]

Questionnaire	Components
Flow Questionnaire [23]	Clear goals High concentration Reduced self-consciousness Distorted sense of time Direct and immediate feedback Balance between ability level and challenge A sense of personal control Intrinsically rewarding activity
Presence Questionnaire [116]	Control factor Sensory factor Distraction Realism factor
IEQ [51]	Emotional involvement Cognitive involvement Real world dissociation Challenge Control
GameFlow Questionnaire [103]	Concentration A sense of challenge Player skills Control Clear goals Feedback Social interaction Immersion
GEQ [14]	Absorption Autonomy Relatedness Presence (immersion) Competence
PENS [96]	Autonomy Relatedness Presence (immersion)
Social Presence in Gaming Questionnaire (SPGQ) [25]	Psychological involvement (empathy) Psychological involvement (negative feelings) Behavioral engagement

B Expanded Summary, Conclusion, and Future Work

In this section, we expand on the conclusions and future work that was presented in the main article. We discuss, in greater detail, the conceptual framework of player engagement, modalities, and predictors of player engagement, establishing the ground truth of engagement, and state-of-the-art estimation models and future directions.

Conceptual Framework of Player Engagement. Conceptually, player engagement spans cognitive, affective, and behavioral dimensions. The cognitive aspect is triggered by audiovisual stimulation, manifesting in heightened concentration and reduced spatio-temporal awareness. The *fun* experienced during gameplay fosters emotional attachment, reinforcing the desire to continue playing. However, this fun is influenced by factors like the balance between game difficulty and player skill, goal clarity, and feedback availability, all of which underpin player engagement. Engagement both influences and is influenced by the player's interaction with the game. For instance, a player may begin a game highly engaged due to its relevance, promotional anticipation, or curiosity. A player consistently experiencing high engagement levels is more likely to continue playing and recommend the game to others. This conceptual complexity directly impacts the measurement of player engagement, particularly in defining it. For example, continuation desire, or conation, can be measured by the number of attempts at a game level [87]. The variety in engagement questionnaires stems from disagreements on the definition of engagement, leading to various questionnaires (e.g., GEQ [14], IEQ [51]) that prompt players to report different gameplay experience aspects, each based on different engagement theories. While these approaches offer a nuanced understanding, the commonality among definitions is the agreement on the cognitive, affective, and behavioral nature of engagement. Therefore, we recommend future research ensures these dimensions are covered in any adopted definition of player engagement.

Modalities and Predictors of Player Engagement. While player engagement indicators can be extracted from various modalities, their reliability and practicality vary. Physiological signals, such as cardiovascular metrics and body temperature, require a stronger theoretical foundation and robust methodology. Since smartwatches are primarily marketed for health monitoring, there's limited interest in using them for game research. A framework is needed to integrate physiological signals from ubiquitous wearable devices with gaming platforms seamlessly. Any modality requiring specialized devices is generally limited to exploratory studies and less suitable for widespread consumer use. For instance, while EEG can provide valuable insights into player engagement, consumer EEG devices are far less ubiquitous than earbuds or smartwatches. However, eye-trackers embedded in increasingly common virtual reality devices offer a vital opportunity to quantify engagement in VR games, especially in areas like immersion and presence. The relatively easy access to facial input, combined with the maturity of facial expression research, has led to a growing trend of using player facial footage to quantify engagement. Although practical, this method requires further validation, particularly in its robustness across various input configurations like frame rate, resolution, and lighting conditions. Additionally, while physiological and neurological features focus directly on the player's state, facial input reflects behavioral responses, which may vary based on game genre, pre-game state, and player personality. Engagement frameworks involving facial input must also address privacy concerns and, in mobile games, battery life. While promising in practicality, this approach is still in its early stages regarding performance validation. Game telemetry, gameplay footage, and user-input data are among the most unobtrusive methods and have shown promise in quantifying player engagement [83]. For systems like cloud gaming, user inputs, and game streams are already collected. This approach requires developers to implement APIs to collect and share telemetry data with gaming systems, as seen in games like PUBG, to quantify engagement. However, the generalizability of these methods across different games and mechanics remains unproven. Industry practices emphasize practical, scalable approaches to engagement measurement, focusing on metrics like DAU/MAU ratios, session length, and retention rates [44, 58]. While academic research explores sophisticated physiological and neurological measurements, industry solutions prioritize readily available data sources and actionable insights [1]. The industry's focus on event-based tracking and custom engagement triggers [45] suggests opportunities for academic research to develop more practical, implementation-focused engagement estimation methods.

Establishing the Ground Truth of Engagement. Establishing the ground truth of player engagement measurement is arguably the most challenging aspect of the problem due to its subjective, multidimensional nature and continuous process. The subjective and multidimensional aspects typically require comprehensive assessments involving questionnaires. However, the continuous nature of engagement means that interrupting gameplay may interfere with the player experience. Previous approaches have circumvented this issue by using third-person observation, time-continuous annotation of recent playthroughs, or proxies such as conation, operationalized as the number of attempts. Third-person observations are prone to bias, as different observers, regardless of experience, cannot objectively capture the player's experience, leading to inconsistent measurements. This is especially relevant in player engagement measurement, as opposed to tasks like emotion recognition, where task-related signs (e.g., sobbing indicating sorrow) are relatively easier to observe. Post-game time-continuous annotations merge the self-reported aspects of questionnaires with the real-time nature of continuous annotations to capture experience fluctuations. However, such annotations are susceptible to recall bias due to the gap between the experience and the annotation [41]. Proxy concepts like conation, though practical, are too narrow to capture the full nuances of player engagement. They are more suited to exploratory studies with large player sets, but their conceptual scope limits their validation capacity. An optimal validation method for player engagement should be continuous and self-reported, with no interruption and no gap between experience and annotation. While this is practically infeasible, combining the aforementioned approaches helps mitigate the limitations.

State-of-the-Art Estimation Models and Future Directions. Current player engagement estimation models span a wide range of approaches, from face-based methods [20, 79, 87] to physiological and behavioral data-based models [35, 71], and advanced techniques like EEG-based indices [94], Bayesian hierarchical models [97], and multimodal analysis [83]. While these models show promise in specific contexts, they face common limitations: limited sample sizes, game-specific applicability, and challenges in capturing engagement's temporal dynamics. Face-based methods struggle with individual expressiveness variations, physiological approaches grapple with data integration, and advanced techniques often lack real-time applicability. To advance the field, future research should focus on developing more generalizable models that can function across diverse gaming contexts and player demographics. Emphasis should be placed on multimodal approaches that combine various data sources (e.g., facial expressions, physiological signals, and gameplay data) to provide a more comprehensive understanding of engagement. Additionally, researchers should explore methods to capture fine-grained temporal dynamics of engagement without disrupting gameplay. Industry experts should consider implementing adaptive systems that can adjust game difficulty or content based on real-time engagement estimates, enhancing player experience. Collaboration between academia and industry could accelerate progress by providing access to larger, more diverse datasets and real-world testing environments.

Received 29 August 2024; revised 27 February 2025; accepted 2 March 2025