A Framework for the Assessment
of Enjoyment in Video Games

Elizabeth Matthews[1(✉)], Geoffrey Matthews[2], and Juan E. Gilbert[1]

[1] Computer & Information Science & Engineering Department,
University of Florida, Gainesville, USA
`lizmatthews@ufl.edu`
[2] Computer Science Department, Western Washington University,
Bellingham, USA

Abstract. We2 present an overview of current research in the measurement of enjoyment in video games. We attempt to systematize the disparate approaches found with the vocabulary and constructs of quasi-experimentation. In each area, we try to make some recommendations for improvement, and find some areas in need of further research.

Keywords: Games · Enjoyment · Flow · Quasi-experiment · Causality

1 Introduction

We propose here a framework to help systematize the study of enjoyment in video games. The importance of video games as a business, as a social problem, and as a psychological tool is difficult to exaggerate, so pervasive have they become today. Perhaps second only to the internet itself, in terms of a modern technology that has changed our world.

The purpose of video games, put simply, is to enjoy them. There are, of course, many elaborations of this possible, and many caveats in its application, but as a touchstone construct this serves well. Accordingly, in order to assess how well video games serve this purpose, we need a common approach and vocabulary for the scientific study of enjoyment in video games. Progress in the measurement of enjoyment in video games, and understanding of its causes, will serve well in designing better video games, in enhancing their effectiveness as serious games, and will also have substantial economic utility in the games industry.

We take it as given that the goal of scientific experiments is the determination of cause and effect. Since our subject of study is the assessment of enjoyment in video games, we phrase this as the study of the cause/effect relation between playing video games and the enjoyment of the experience. If we can define and measure this relation scientifically, our experiments might help in the answer to such salient questions as:

– Does playing a video game cause enjoyment in some subjects more than others?

– Which kinds of video games augment which types of enjoyment? For which types of subjects?
– Which features of video games are most effective at enhancing enjoyment? Which should be added, or removed, to shape the experience?
– How is the enjoyment of video games different from the enjoyment of movies or literature?
– Are video games effective at treating depression?
– Which kinds of people are most affected by video games? Least affected?
– What are effective means of treating video game addiction?
– How can we make video games more appealing to a given market sector?

There is a large literature on the measurement of enjoyment *per se*, and a good portion of this is focused specifically on enjoyment in video games. We believe that the time has come for a roadmap to this literature, a context in which the contributions of each experiment can be seen as contributing to a whole. This paper is our attempt to begin a dialogue on this task, and perhaps approach some kind of consensus as to vocabulary, categories, and the like.

2   The Structure of Experiments and Quasi-Experiments

The scientific validity of the experiments establishing the relationship between video games and enjoyment is critical for answering these questions and a host of others. We would do well, at this early stage in the development of these scientific studies, to make sure that we are not working at cross-purposes, that one study can be reliably compared with another, and that progress as a whole will be steady.

To this end we propose a framework that seems to fit well with ongoing work, identifies problems in designs, and also suggests avenues for further research. The framework we advocate here is based on the landmark work on experimental and quasi-experimental design, Shadish et al. [1]. The factors involved in an experimental or quasi-experimental design are *units*, *treatments*, *observations*, and *settings*. We make a few remarks here on each of these in the context of video games.

Units. The unit of study, as in most psychological experimentation, is the person. Given that gaming is a self-selected activity, and the response construct is enjoyment, strict random sampling would be counterproductive. However, given that gaming enthusiasm has a complex metric with broad spectra indicates that other sampling designs (such as stratified sampling or regression discontinuity) show much promise. Work still needs to be done in the taxonomy of gamers and their responses to game feature modifications so that future experiments can take advantage of these profiles.

Treatment variables. What do we change and manipulate in order to assess its impact on enjoyment? In some experiments the treatment variable may be the simple presence or absence of games (or the extent to which games appear). Studies of this type are useful, for example, in studying the efficacy

of psychiatric therapy using video games and contrasting it with alternative therapies that do not use video games. In other experiments, the treatment variable can be the presence or absence in a game of various features, such as virtual reality, physical interactivity, multiplayer modes, or procedural generation. Studies of this type are useful in the cost/benefit analyses of game companies, and this type of study is far more common than studies of the efficacy of the simple presence or absence of games. We will devote the majority of our attention to this kind of study.

When we consider the possible treatment variables for a study of enjoyment in video games, we come up with such things as how much action is in the game, how much puzzle solving, the theme of the game, or whether procedural generation is used in the game. These factors are usually studied under the rubric of game design decisions, as in Schell's work [2], usually with the implicit understanding that good design is driven by making the game more enjoyable. Dimensions along which profitable treatment variables can be designed include causal *vs.* serious, physics *vs.* logic, single-player *vs.* multiplayer, action *vs.* puzzle, and procedurally generated *vs.* static. Further, these can often be productively combined into two and higher dimensional maps.

Observations, or response variables. What do we measure when we attempt to estimate the level of enjoyment? Such measures usually take the form of subjective measurements, such as those provided by questionnaires, and objective measurements, which are provided by capturing physical data about a subject. Objective measurements are further subdivided into physiological measurements, such as heart rate and breathing, and behavioral measurements, such as facial expressions, mouse movements, and clicking rate. Each of these has seen widespread use in the literature, however the interrelations among them raise issues for construct validity which we address in Sect. 3. Construct validity deals with the relationship between measurable response variables and higher level constructs, such as the experience of "flow," or, indeed the experience of "enjoyment" as a response to the treatment. In what follows, when the relationship between measurable responses and higher level constructs is not specifically being addressed, we will sometimes refer to the higher-level constructs themselves as response variables.

Settings. The normal setting for a game is without context. The game is played for its own intrinsic merit and little else contributes to the overall experience. However, there are many important contexts in which the game plays a part in a larger scenario. The game may be played for internal reasons, (for example, to relax, to escape, or even because of addiction) or external reasons (for example, to evaluate a subject's psychological state, to evaluate their intelligence, or even to treat a problem). This context is important and may contribute a confounding factor to the enjoyment of the game when the subject is aware of it.

The utility of these categories can be seen immediately in identifying and classifying threats to the validity of causal inference. A threat is the potentiality

of making an incorrect inference to cause and effect. Following Shadish *et al.* [1] we can identify four broad categories, and our subsequent discussion is organized accordingly:

Statistical conclusion validity. Are the statistical tests and inferences valid, given the data? This is a straightforward statistical question and, while important, will not be discussed in this paper as there exists ample literature addressing the various statistical tests.

Construct validity. We measure response variables, for example heart rate, or reported satisfaction on a questionnaire. What confidence do we have that the measurements we actually take reflect what we are trying to measure, namely, enjoyment? Further, mid-level constructs, between the response variables and the construct of enjoyment itself, dominate the literature on game enjoyment. Examples of such mid-level constructs include game flow, motivational states (including needs satisfaction), emotional states, and engagement. We find that this is a particularly large and difficult area in game enjoyment studies and will spend some time cataloging approaches and possible threats in Sect. 3.

Internal validity. When we find that the response variables do indeed covary with the treatment variables, under what circumstances are we entitled to conclude that there is a causal relation? Under the assumption of random treatment groups, this problem has well-understood statistical gounds. However, in game studies, given the complex taxonomy of gamers, completely random selection would be problematic. Without random selection (a situation identified as "quasi-experimentation" in Shadesh *et al.* [1]) there are many situations, such as regression to the mean, that are well understood threats to internal validity. The primary approach to dealing with these threats is to attempt to rule out, as exhaustively as possible, other possible causes. We address briefly the specialization of these threats and their solutions to game enjoyment studies in Sect. 4.

External validity. Does the cause-effect relationship found in a given experiment generalize to other persons, settings, treatment variables or response variables? These questions are obviously relevant to marketing in the games industry, but also to serious games, for example, the use of games for education or psychological therapy. We discuss these issues briefly in Sect. 5, and call for further investigation into the proper taxonomy of gamers. We believe that this could be a highly fruitful field for the application of machine learning, for example, clustering algorithms, in the near future.

3   Threats to Construct Validity in Game Enjoyment Studies

Construct validity addresses the question of the relation between the response variables and the construct under study. In this case the construct is enjoyment, and so we are engaged with the questions: "What is enjoyment, and how do

we measure it?" This is a large topic, and we will spend the majority of this paper summarizing the answers found in the literature. Enjoyment *per se*, like all abstract categories, is difficult to define, but as it is a nontechnical term in widespread use, we feel it not necessary to define, but as a guide we favor the definition of Merkler [3]: the positive cognitive and affective appraisal of game experience.

Most research on video game enjoyment, however, uses more specific constructs, or what we call "mid-level constructs," in between measurable response variables (*e.g.* heart rate or questionnaire responses), and the construct of interest, *viz.* enjoyment. Some of the most important mid-level constructs include: GameFlow, Motivational States, Emotional States, Needs Satisfaction, and Engagement. Each of these, considered as response constructs, carries its own concerns in regard to construct validity. There are three broad categories of questions that need answers in any study involving them:

1. How do measurable response variables relate to the mid-level constructs?
2. How do mid-level constructs relate to the overarching construct of enjoyment?
3. How do the mid-level constructs relate to each other?

We address several popular mid-level constructs in the following subsections, and we conclude each section with some remarks on the relative utility of subjective and objective measures to each of them.

3.1   GameFlow

Flow, first proposed by Csikszentmihalyi and Csikszentmihalyi [4], is the idea that people find genuine satisfaction in a state of consciousness, achieved by tailor-fitting the subject matter to each individual's skills, being neither too demanding nor too easy. GameFlow applies Flow to games, as proposed by Sweetser and Wyeth [5]. Currently it is one of the most commonly used constructs in game enjoyment studies. A strong benefit of the GameFlow definition of enjoyment is that the model has its roots in theories formed in other disciplines. Interactive digital media is a relatively new phenomenon, and being able to build on established and tested theories to create a new method lends a strength to the method.

The first caveat that arises in considering GameFlow as a response variable is that GameFlow (and Flow in general) is defined as a mixture of both objective factors about the difficulty level of a game, and subjective factors regarding the player's experience of the game. Objective factors, for example, include: (1) the game presents a task that can be completed and (2) the task provides immediate feedback. Subjective factors, for example, include: (1) the task has clear goals, (2) there is a sense of control over actions, (3) a deep but effortless involvement, (4) concern for self disappears, but sense of self emerges stronger afterwards, and (5) the sense of the duration of time is altered. Only the subjective aspects of GameFlow make reasonable response variables.

Sweetser and Wyeth [5] adapt this checklist from the original Flow description into the following categories: Concentration, Challenge, Skills, Control,

Clear Goals, Feedback, Immersion, and Social Interaction. Each category covers a variety of the items from Flow's definition, and each category covers a multitude of criteria. Challenge, Clear Goals, and Feedback would seem to be objective features of the game, while Concentration and Immersion are subjective experiences, while the rest would seem to be a mix of the two. Again, experimental designs must make clear when evaluating GameFlow specifically as a measure of enjoyment (and not, for example, as a measure of the quality of the game), which subjective features of GameFlow are the constructs of interest.

Flow adapted to games is also proposed by Chen [6], who states that game players have a wide range of skill, so a single game design experience can not fully guarantee that all users will stay within the zone. This is clearly a testable hypothesis if we are careful to design the experiment around measurable response variables.

A second difficulty in using GameFlow as a response variable is its observability problem. Much like Heisenberg's uncertainty principle in quantum mechanics, to observe it is to destroy it. Nacke and Lindley [7,8] used the Game Experience Questionnaire (GEQ; IJsselsteijn et al. [9]) and physiological measurements to measure GameFlow and immersion in first person shooter type games. In the Nacke and Lindley 2010 [8] study it was discovered the GEQ was unable to measure immersion and boredom, but was able to measure flow. Nacke and Lindley provide guidelines in their paper for design criteria for designing around GameFlow. Weber et al. [10] recommend using unobtrusive physiological measurements (specifically, fMRI) for measuring the state of flow so that it does not disrupt the experimental state.

A benefit to the GameFlow model is that it can be extended to cover various subsets of playtesting, either game or user type specific, such as the motor-impaired users (MIU)-GameFlow model [11], the GameFlow model for Pervasive Games (PGF Model) [12,13], or the EGameFlow model [14]. The EGameFlow model is an adaptation of the GameFlow model to apply to learning games specifically proposed by Fu et al. [14]. The EGameFlow model follows the original GameFlow categories with some modifications, such as the addition of the category of "Knowledge Improvement," and converts the criteria into Likert-scalable statements.

One of the most difficult problems in using GameFlow as a response variable is the confusion between the "Goldilocks" definition of GameFlow (not too hard, not too easy), and the definition of GameFlow as a psychological experience: being "in the zone," losing track of time, *etc.* The level of difficulty, relative to a player, is fairly straightforward to measure. The intensity of the flow experience is extremely difficult to measure. A lot of research assumes that measuring one is correlated with the other, but this is a hypothesis that, in our view, remains unconfirmed. Future work in this area should try to firmly establish the connection rather than rely on an implicit understanding of the words "Game Flow."

Tools need to be developed specifically for testing of GameFlow. While there are previously developed scales for entertainment that can serve as a starting

reference, these scales do not always work for GameFlow, and in fact may be confusing or conflicting. Procci and Bowers [15] examined flow and immersion within games using the Dispositional Flow State Scale (DFS-2) and the Immersive Tendencies Questionnaire (ITQ). Procci and Bowers findings were that the two scales were not overlapping, despite similar items being measured on each questionnaire, and therefore cannot be used interchangeably.

Most studies use some form of questionnaire based measurements when accounting for GameFlow, including the models EGameFlow [14] and the GEQ [9]. In this case, Flow and GameFlow occur when the user achieves an "in the zone" state, and using self-reporting measurement tools brings the user out of one mindset and into another for answering questions. Future work should seriously consider unobtrusive physiological measurements, following the recommendation of Weber et al. [10].

3.2   Motivational States

The consideration of Motivational States allows the definition of enjoyment to encompass the "Pre-Game" phase presented in the Integrated Model of Player Experience [16]. By examining what may drive a player, one gets a fuller picture of the player's intentions and how satisfying those motivational states increases player enjoyment. In this case, the enjoyment is not strictly caused by playing a game, but rather by the history of the player's interactions with games.

Motivational States are important because enjoyment effects may not be entirely dependent on flow states. Kaye [17] found that motivational states did influence the enjoyment of a game as well as elements of flow theory. Kaye proposes a framework for modeling motivational states and their effect on game type selection and enjoyment. The model has external factors providing both player and game-type motivations into the type of game selected. Motivational states allow the entire gaming experience to be considered for a more comprehensive understanding of the enjoyment process.

Motivational states can be categorized into two types: extrinsic and intrinsic [18]. Extrinsic motivation comes from external factors, such as monetary gain upon completion of a task. Intrinsic motivation comes from the task itself and a particular person's own goals and desires. Within these two categories, three formats are possible: pleasant experiences, ethical motivations, and goal setting/achievement. Cota et al. [19] confirm that several motivational aspects for elderly players fell under the appropriate categories of intrinsic/extrinsic and player preference/ethical motivations/goal setting and achievement.

Some tools are available for motivational states measurements. For example, the Attention, Relevance, Confidence, and Satisfaction model (ARCS) presented by Keller [20]. The ARCS model was developed further into the Instructional Materials Motivation Survey by Keller [21].

Derbali and Frasson [22] and Derbali et al. [23] used the Instructional Material Motivation Scale (IMMS) [21] questionnaire to correlate between objective physiological measurements and motivational states. Theta waves in the frontal regions of the brain and motivation were positively correlated, high-beta waves

in the left-center region were a significant predictor for high level of motivation, and skin conductance was a significant predictor for motivation. The correlation between these subjective and objective response variables recommends them for experimental design, and helps answer some questions raised by Ghergulescu and Muntean [24], who suggest that using subjective self-reporting measures only reflects motivational states at the time of the questionnaire rather than the gamer's actual motivational states throughout the game play session.

Needs Satisfaction. Needs Satisfaction as a measure of enjoyment is based on Self Determination Theory (SDT) [25]: the concept that enjoyable actions satisfy a base need for the subject. This is clearly a motivational approach, but is specific enough that we give it a separate section here.

The Player Experience of Need Satisfaction model (PENS) was developed by Rigby and Ryan [26] and applies the three needs of Competence, Autonomy, and Relatedness to gaming experiences. The PENS was also used by Tamborini et al. [27], and Neys et al. [28] demonstrated that Autonomy, Competence, and Relatedness can be successfully measured by the PENS.

The Situational Motivation Scale (SIMS) [29], which is also based in SDT, can measure regulation modes of Intrinsic Motivation, Identified Regulation, External Regulation, and Amotivation [28].

Tamborini et al. [27] refine the Needs Satisfaction model of enjoyment, distinguishing hedonic and nonhedonic needs. Hedonic needs are arousal and absorption, and nonhedonic needs are competence and autonomy. They again used the PENS to measure the needs satisfaction in the two categories of autonomy and competence. The remaining two categories of arousal and absorption were each measured with three-item Likert-type scales Tamborini et al. [27] created. Measurements showed that low interactivity creates low arousal, and both types of needs were statistically significant in positive correlation with self-reported/subjective enjoyment.

The Intrinsic Motivation Inventory (IMI) [30,31] can be used to measure needs satisfaction. Rieger et al. [32] used the SES questionnaire [33] to measure emotions. Video games are shown to serve mood repair and to help increase positive mood states and to decrease negative mood states by satisfying the needs of participants [32]. In-game success, such as defeating an enemy or scoring a goal, is important to positive moods; however, enjoyment relies more on needs satisfaction than success [32]. The IMI should be compared with measures discussed in Sect. 3.3, on the Emotional States model of enjoyment.

The advantage of the Needs Satisfaction model of enjoyment for the development of reliably measured response variables lies in its specificity. A well established model, the PENS, exists and has been thoroughly tested, as seen with Tamborini et al. [27,34] and Neys et al. [28]. Further, several established questionnaires exist for measurement purposes of needs satisfaction: the IMI [29,31,32], and the SIMS [28]. Due to its basis in SDT there is related work in other disciplines to provide context on the theory.

This specificity of Needs Satisfaction, however, is a double-edged sword. It may help make the response variables more concrete, but their generalization to enjoyment in general is accordingly more suspect. Further, minimal work has been done on the correlation between objective measurements and needs satisfaction, forcing researchers to rely on subjective measurements. Further research is needed in this area.

3.3   Emotional States

Similar to the Motivational States model is the Emotional States model. Madeira et al. [35] present an examination of psychological game theory. They also state that both subjective and objective measurements can be used when measuring emotional states. Haag et al. [36] recommend multiple approaches to measure emotional states.

When measuring emotional states, two kinds of emotional value are typically measured: Arousal (emotional strength of the content) and Valence (the positive or negative consideration rating of the content). Bio-sensors [36] and brain activity/electroencephalograms (EEG) [37] have been shown to be reliable measuring tools.

A combinational system for measurement is proposed by Kivikangas et al. [38]. Participants self-report experiences of game events via review of automatically created video clips and questionnaires about the events. This self-reporting method was also supported by physiological measurements. Another combination system is Biometric Storyboards [39].

Self-reporting measurements and questionnaires, such as the Positive and Negative Affect Schedule (PANAS) [40], have been used to capture emotional states. Studies of this type have shown that emotional states vary greatly for individuals [41] and emotions are freely felt during solitary play [42].

The advantage of Emotional States over Motivational states is that they are somewhat more amenable to objective measurement. Both self-reporting, such as the PANAS, and physiological, such as the approach developed by Kivikangas et al. [38], are shown to be fairly reliable measurements for emotional states. The variety of available measurements for this definition of enjoyment allows for diverse study types. In addition, the two dimensional nature (arousal and valence) of Emotional States allows more sophisticated analysis of the response.

However, Emotional States are more difficult to make commensurable between subjects than Motivational States, for emotional states are different between individuals [41]. This in turn will require sophisticated analysis of the data to guard against threats to statistical validity.

3.4   Engagement

Engagement is frequently used as a catch-all construct for describing enjoyment in video games, and can encompass other constructs such as immersion, enjoyment, presence, flow, and arousal [43]. This makes it particularly difficult to use

as a response variable, in that it is incumbent on the researcher to nail down exactly what is being measured. We summarize here some attempts to do this.

Silpasuwanchai et al. [44] make an attempt to clarify engagement, and separate engagement into three categories: emotional, behavioral, and cognitive. Emotional engagement is the valence, arousal, and endurance of the evoked affective state, similar to emotional states (discussed in Sect. 3.3). Behavioral engagement is how a participant's behavior may change based on their engaged status, and cognitive engagement is when stimulus creates a situation where a participant is mentally stimulated into higher-level thinking.

Another framework to measure engagement is based in Continuation Desire (CD; Schoenau-Fog et al. [45]). The model has many items, such as emotional engagement or intrinsic attention focus, all linked to CD. Continuation Desire is the desire or willingness to continue an experience, and can be used as a metric to measure the quality of an interactive story experience. This has aspects of both the Motivational States and the Emotional States approaches (Sects. 3.2 and 3.3).

Another approach to engagement measures is the Traces model [46–48]. Bouvier et al. combine the motivational concept of SDT, Activity Theory, and Trace Theory to explain game engagement. Based on an SDT sourced definition of engagement the four kinds of engaged behaviors can be defined: (a) environment-directed (exploration and modding), (b) social-directed (expanding social network or sharing with others), (c) self-directed (character customization or story creation), and (d) action-directed (mastering a game skill or elaborating a strategy).

Trace theory considers the behavior of a gamer as a sequence of actions taken, such as mouse clicks or keyboard input. At the base of the framework are observed events, called obsels, which contain the type of event, a timestamp, and a set of contextual information. A trace is a set of obsels that may be connected. In testing their theory, Bouvier et al. [47] had highly accurate results, with an accuracy rating of 91.67% for engagement prediction, 80% for prediction of social-engagement, and 100% for both action-engagement and environment-engagement.

Marsh and Nardi [49] suggest an activity-based approach to engagement, with additional focus on motivations per objective. The framework proposed is to consider a sphere of engagement through motive in activity. Actions that share a motive are contained within a sphere of engagement. They provide a flexible framework for future analysis/design of interactive digital media.

Procci [50] provides an examination of the Revised Game Engagement Model (R-GEM) based on immersion, involvement, presence, and flow. The model categorizes immersion and involvement as low-level game engagement, and presence and flow as high-level engagement Results from a study showed that the model still needed work but generally showed reliable factors [50]. Emotion and engagement are both biological and subjective constructs, and a combination of physiological and self-reporting methods are recommended [51].

Another tiered approach to engagement is on a scale from low to high levels of engagement, moving from immersion to presence, then flow, and finally absorption [52]. The research examined a "low-level" game engagement score (immersion plus presence subscales) versus a "high-level" game engagement score (flow plus absorption subscales). A low-level game engagement score was a strong predictor of high-level game engagement score, emphasizing the idea that game engagement is a scale. Procci et al. [52] examined several influencing factors, but the only significant effect they found was that age had a negative effect on high levels of engagement and decreased the relationship between low and high engagement prediction.

Behavioral cues can be used to measure engagement based on flow concepts [53,54]. Behavioral measurements can be obtained without intrusive measuring equipment. Riemer and Schrader [54] measured Behavioral Engagement in High Relevance phases (BEHR) and Behavioral Engagement in Low Relevance phases (BELR) in an educational gaming setting. High relevance phases are moments where a game user may self-reflect or exhibit self-monitoring behavior. Low relevance phases are behaviors exhibited with low relevance to the educational objective. They reported that only self-monitoring affected mental model development in serious games; behavioral engagement had no effect.

One advantage to defining enjoyment as engagement is its encompassing nature. Engagement typically contains aspects from other well-established enjoyment definitions of immersion, involvement, presence, and flow [50]. The overarching nature of enjoyment is also reflected in engagement, as many aspects make up what is enjoyable about an experience. Using this definition can allow a broader approach to a research idea on enjoyment.

EEG measurements have been proven to accurately record engagement [55,56]. McMahan et al. found that EEG off-the-shelf modules (such as the Emotiv) can reliably be used to measure gamer's engagement during game play, specifically relating to player events that occurred within the game (death, normal play).

The all-encompassing nature of engagement is also a vulnerability. Due to the diverse nature of engagement, using the term does little to help clarify exactly what the research is measuring. It is best used, therefore, in conjunction with additional specific definitions to define the research focus.

4   Threats to Internal Validity in Game Enjoyment Studies

The establishment of cause and effect is subtle and should be given careful consideration in the design of experiments. However, as noted by Shadish *et al.* [1], at a minimum we need to establish three things: First, causes *precede* effects in time. Second, causes *covary* with their effects. Finally, alternative explanations of the effects are *implausible.* The first two are fairly obvious and due diligence is usually paid to them in the design of experiments. It is the third condition that is the most difficult to address, and causes the greatest amount of confusion and

error in experimental design. It is essentially the question of *internal validity*: does the observed covariance warrant the conclusion of a causal connection? For this reason we will be particularly concerned with designing our experiments on enjoyment in video games to facilitate this effort.

Ideally, random assignment to treatment and control groups provides assurances of internal validity. However, the complex and as yet poorly understood taxonomy of gamers makes true random assignment difficult. It is unlikely that all types of gamers will be represented equally, unless the sample size is huge, and it is still unclear how to stratify subjects in this respect. As a result, alternatives to random assignment take precedent. In particular, the effort to *rule out alternative explanations* of the covariance seen in experiments must be part of the experimental design.

Without random selection and control groups we are left with what is termed quasi-experimentation. Designs that lack control groups and pretest observations are discussed in general in Shadesh *et al.* [1], Chap. 4. In many simple studies of video game enjoyment, such as the work proposed by Korn *et al.* [57], the designs are variations of the *removed-treatment design*. Basically, the subjects play two different kinds of games, say, with and without a certain feature, and their enjoyment levels are then observed. This can be diagrammed, for example, as one of the two designs in Fig. 1 where time moves left to right, the Os represent different observations of the response variables, the X represents treatment (for example playing a game with an added feature), and the $\bar{X}$ represents absence of treatment (playing the game without the added feature). If the response variables are measured with surveys, they would come immediately after playing the game; if they were physiological measurements, they would come simultaneously with game play.
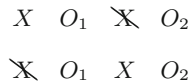
$$X \quad O_1 \quad \bar{X} \quad O_2$$

$$\bar{X} \quad O_1 \quad X \quad O_2$$

Fig. 1. Removed-treatment experimental design.

One problem with this design is the difficulty of accounting for novelty effects, fatigue effects, practice effects, carryover effects, order effects, *etc.*, many of which are quite common in video game studies (*e.g.* any change in a game will generally seem more interesting than playing the old version, at least at first). To address this the *repeated treatment design* seems most apt to these situations. It is pictured in Fig. 2. The idea here is that if, for example, we expect the treatment to increase the value of the response variable, then we would confirm this expectation by observing an increase between O_1 and O_2, a decrease between O_2 and O_3, another increase between O_3 and O_4, *etc.* The pattern can obviously be continued indefinitely. However, in game studies, fatigue effects are also problematic (any game will generally seem less and less interesting as time goes by).

$$O_1 \quad X \quad O_2 \quad \cancel{X} \quad O_3 \quad X \quad O_4 \quad \cancel{X} \quad O_5 \quad \dots$$

Fig. 2. Repeated treatment design

To address both of these, we suggest a *cohort repeated treatment design* in which cohorts of subjects participate in subsections of a repeated design. For example, with three administrations of the treatment, each of the eight possible patterns could be followed by a different group of the subjects, as illustrated in Fig. 3. If there are more treatments than two (more than just "presence" or "absence"), a Latin square [58] or other disciplined selection of sequences can be used to avoid the exponential explosion of possibilities. Sequences longer than three could also be considered, so long as they do not introduce fatigue effects (which could be tested for using groups 1 and 8 from the design in Fig. 3).

$$O_1 \quad X \quad O_2 \quad X \quad O_3 \quad X \quad O_4 \tag{1}$$
$$O_1 \quad \cancel{X} \quad O_2 \quad X \quad O_3 \quad X \quad O_4 \tag{2}$$
$$O_1 \quad X \quad O_2 \quad \cancel{X} \quad O_3 \quad X \quad O_4 \tag{3}$$
$$O_1 \quad X \quad O_2 \quad X \quad O_3 \quad \cancel{X} \quad O_4 \tag{4}$$
$$\dots$$
$$O_1 \quad \cancel{X} \quad O_2 \quad \cancel{X} \quad O_3 \quad \cancel{X} \quad O_4 \tag{8}$$

Fig. 3. Cohort repeated treatment design.

The central idea here is to introduce a complex pattern of treatments. The more randomly distributed, over time, the treatments are, the more they should resemble a truly randomized experiment. If we observe a consistent pattern in the responses, well-correlated with the treatments, this substantially decreases the likelihood of an alternative explanation, and increases the probability that the treatment had a genuine causal effect.

5   Threats to External Validity in Game Enjoyment Studies

Here we get into questions that are surely in the hearts of every marketing department: how do we sell games to customers who haven't played games before? When will our measurements of gamer enjoyment generalize to a population that is not characteristically exposed to games?

Considerations as to the generalization of settings are also important for such areas as intelligent games, psychotherapeutic games, *etc.* When can we be sure that our findings about game enjoyment, and hence greater attraction to the gaming experience, can be generalized to contexts where, for example, games are not played for their own sake? What considerations need be addressed in the

design of experiments on game enjoyment to give credence to their generalization over settings?

Little research has been done addressing this essential question. One thing that could be given more consideration in designing experiments is the use of multiple, independent response variables, so far as this is feasible. Our remarks above, in Sect. 3, on such constructs as GameFlow, Motivational States, Emotional States, and Engagement, are relevant, here. These constructs are, of course, related in complex ways, but also distinct. The more we can consider each of them as an independent measure from the others, the more validity they will give to our studies, and consequently greater security against threats to external validity. To this end, both subjective and objective response variables should be as focused as possible to measure distinct reactions, distinct components of enjoyment.

6   Conclusion

We believe that the study of enjoyment in video games is well begun. We maintain here that using the framework and vocabulary of quasi-experimentation will enable better understanding of the results, and more synergy between researchers. Further, in addressing various threats to validity, new avenues for research are suggested.

References

1. Shadish, W., Cook, T., Campbell, D.: Experimental and Quasi-experimental Designs for Generalized Causal Inference. Houghton Mifflin, Boston (2002)
2. Schell, J.: The Art of Game Design: A Book of Lenses. Morgan Kaufmann Publishers Inc., San Francisco (2008)
3. Mekler, E.D., Bopp, J.A., Tuch, A.N., Opwis, K.: A systematic review of quantitative studies on the enjoyment of digital entertainment games. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 927–936. ACM (2014)
4. Csikszentmihalyi, M., Csikszentmihalyi, I.S.: Optimal Experience: Psychological Studies of Flow in Consciousness. Cambridge University Press, Cambridge (1992)
5. Sweetser, P., Wyeth, P.: Gameflow: a model for evaluating player enjoyment in games. Comput. Entertain. (CIE) **3**(3), 3–3 (2005)
6. Chen, J.: Flow in games (and everything else). Commun. ACM **50**(4), 31–34 (2007)
7. Nacke, L., Lindley, C.A.: Flow and immersion in first-person shooters: measuring the player's gameplay experience. In: Proceedings of the 2008 Conference on Future Play: Research, Play, Share, pp. 81–88. ACM (2008)
8. Nacke, L.E., Lindley, C.A.: Affective ludology, flow and immersion in a first-person shooter: Measurement of player experience. arXiv preprint arXiv:1004.0248 (2010)
9. IJsselsteijn, W., de Kort, Y., Poels, K.: The game experience questionnaire. Manuscript in Preparation (2008)
10. Weber, R., Tamborini, R., Westcott-Baker, A., Kantor, B.: Theorizing flow and media enjoyment as cognitive synchronization of attentional and reward networks. Commun. Theory **19**(4), 397–422 (2009)

11. Zain, N.H.M., Jaafar, A., Razak, F.H.A.: Enjoyable game design: validation of motor-impaired user gameflow model. Int. J. Comput. Theory Eng. **8**(2), 116 (2016)
12. Jegers, K.: Pervasive game flow: understanding player enjoyment in pervasive gaming. Comput. Entertain. (CIE) **5**(1), 9 (2007)
13. Jegers, K.: Pervasive gameflow.: a validated model of player enjoyment in pervasive gaming. (2007)
14. Fu, F.L., Su, R.C., Yu, S.C.: Egameflow: a scale to measure learners enjoyment of e-learning games. Comput. Educ. **52**(1), 101–112 (2009)
15. Procci, K., Bowers, C.: An examination of flow and immersion in games. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 55, pp. 2183–2187. SAGE Publications (2011)
16. Angelides, M.C., Agius, H.: Know thy player: an integrated model of player experience for digital games research (2014)
17. Kaye, L.K.: Motivations, experiences and outcomes of playing videogames. Ph.D. thesis, University of Central Lancashire (2012)
18. Frey, B.S., Osterloh, M.: Successful Management by Motivation: Balancing Intrinsic and Extrinsic Incentives. Springer, Heidelberg (2002). https://doi.org/10.1007/978-3-662-10132-2
19. Cota, T.T., Ishitani, L., Vieira, N.: Mobile game design for the elderly: a study with focus on the motivation to play. Comput. Hum. Behav. **51**, 96–105 (2015)
20. Keller, J.M.: Motivational Design for Learning and Performance: The ARCS Model Approach. Springer, Boston (2010). https://doi.org/10.1007/978-1-4419-1250-3
21. Keller, J.M.: IMMS: instructional materials motivation survey. Florida State University (1987)
22. Derbali, L., Frasson, C.: Prediction of players motivational states using electrophysiological measures during serious game play. In: 2010 10th IEEE International Conference on Advanced Learning Technologies, pp. 498–502. IEEE (2010)
23. Derbali, L., Chalfoun, P., Frasson, C.: Assessment of learners' attention while overcoming errors and obstacles: an empirical study. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) AIED 2011. LNCS (LNAI), vol. 6738, pp. 39–46. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21869-9_8
24. Ghergulescu, I., Muntean, C.H.: A novel sensor-based methodology for learner's motivation analysis in game-based learning. Interact. Comput. **26**(4), 305–320 (2014)
25. Deci, E.L., Eghrari, H., Patrick, B.C., Leone, D.R.: Facilitating internalization: the self-determination theory perspective. J. Pers. **62**(1), 119–142 (1994)
26. Rigby, S., Ryan, R.: White paper: the player experience of need satisfaction (pens). Technical report, Immersyve, Inc., September 2007
27. Tamborini, R., Bowman, N.D., Eden, A., Grizzard, M., Organ, A.: Defining media enjoyment as the satisfaction of intrinsic needs. J. Commun. **60**(4), 758–777 (2010)
28. Neys, J.L., Jansz, J., Tan, E.S.: Exploring persistence in gaming: the role of self-determination and social identity. Comput. Hum. Behav. **37**, 196–209 (2014)
29. Guay, F., Vallerand, R.J., Blanchard, C.: On the assessment of situational intrinsic and extrinsic motivation: the situational motivation scale (SIMS). Motiv. Emot. **24**(3), 175–213 (2000)
30. Ryan, R.M., Deci, E.L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am. Psychol. **55**(1), 68 (2000)
31. Reinecke, L., Tamborini, R., Grizzard, M., Lewis, R., Eden, A., David Bowman, N.: Characterizing mood management as need satisfaction: the effects of intrinsic needs on selective exposure and mood repair. J. Commun. **62**(3), 437–453 (2012)

32. Rieger, D., Wulf, T., Kneer, J., Frischlich, L., Bente, G.: The winner takes it all: the effect of in-game success and need satisfaction on mood repair and enjoyment. Comput. Hum. Behav. **39**, 281–286 (2014)
33. Hampel, R.: Adjektiv-Skalen zur Einschätzung der Stimmung (SES) (1977)
34. Tamborini, R., Grizzard, M., David Bowman, N., Reinecke, L., Lewis, R.J., Eden, A.: Media enjoyment as need satisfaction: the contribution of hedonic and nonhedonic needs. J. Commun. **61**(6), 1025–1042 (2011)
35. Madeira, F., Arriaga, P., Adrião, J., Lopes, R., Esteves, F., et al.: Emotional gaming. Psychol. Gaming **1**, 12–29 (2013)
36. Haag, A., Goronzy, S., Schaich, P., Williams, J.: Emotion recognition using biosensors: first steps towards an automatic system. In: André, E., Dybkjær, L., Minker, W., Heisterkamp, P. (eds.) ADS 2004. LNCS (LNAI), vol. 3068, pp. 36–48. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24842-2_4
37. Horlings, R., Datcu, D., Rothkrantz, L.J.: Emotion recognition using brain activity. In: Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for Ph.D. Students in Computing, p. 6. ACM (2008)
38. Kivikangas, J.M., Nacke, L., Ravaja, N.: Developing a triangulation system for digital game events, observational video, and psychophysiological data to study emotional responses to a virtual character. Entertain. Comput. **2**(1), 11–16 (2011)
39. Mirza-Babaei, P., Nacke, L.E., Gregory, J., Collins, N., Fitzpatrick, G.: How does it play better?: exploring user testing and biometric storyboards in games user research. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1499–1508. ACM (2013)
40. Watson, D., Clark, L.A., Tellegen, A.: Development and validation of brief measures of positive and negative affect: the panas scales. J. Pers. Soc. Psychol. **54**(6), 1063 (1988)
41. Landowska, A., Wróbel, M.R.: Affective reactions to playing digital games. In: 2015 8th International Conference on Human System Interaction (HSI), pp. 264–270. IEEE (2015)
42. Deterding, S.: The joys of absence: emotion, emotion display, and interaction tension in video game play. In: Proceedings of FDG 2015 (2015)
43. Boyle, E.A., Connolly, T.M., Hainey, T., Boyle, J.M.: Engagement in digital entertainment games: a systematic review. Comput. Hum. Behav. **28**(3), 771–780 (2012)
44. Silpasuwanchai, C., Ma, X., Shigemasu, H., Ren, X.: Developing a comprehensive engagement framework of gamification for reflective learning. In: Proceedings of the 2016 ACM Conference on Designing Interactive Systems, pp. 459–472. ACM (2016)
45. Schoenau-Fog, H., Louchart, S., Lim, T., Soto-Sanfiel, M.T.: Narrative engagement in games-a continuation desire perspective. In: FDG, pp. 384–387 (2013)
46. Bouvier, P., Sehaba, K., Lavoué, E., George, S.: Using traces to qualify learner's engagement in game-based learning. In: 2013 IEEE 13th International Conference on Advanced Learning Technologies, pp. 432–436. IEEE (2013)
47. Bouvier, P., Sehaba, K., Lavoué, É.: A trace-based approach to identifying users engagement and qualifying their engaged-behaviours in interactive systems: application to a social game. User Model. User-Adap. Inter. **24**(5), 413–451 (2014)

48. Bouvier, P., Lavoué, E., Sehaba, K.: Defining engagement and characterizing engaged-behaviors in digital gaming. Simul. Gaming (2014). https://doi.org/10.1177/1046878114553571
49. Marsh, T., Nardi, B.: Spheres and lenses: activity-based scenario/narrative approach for design and evaluation of entertainment through engagement. In: Pisan, Y., Sgouros, N.M., Marsh, T. (eds.) ICEC 2014. LNCS, vol. 8770, pp. 42–51. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-45212-7_6
50. Procci, K.C.: The subjective gameplay experience: an examination of the revised game engagement model. Ph.D. thesis, University of Central Florida Orlando, Florida (2015)
51. Bardzell, J., Bardzell, S., Pace, T., Karnell, J.: Making user engagement visible: a multimodal strategy for interactive media experience research. In: CHI 2008 Extended Abstracts on Human Factors in Computing Systems, pp. 3663–3668. ACM (2008)
52. Procci, K., James, N., Bowers, C.: The effects of gender, age, and experience on game engagement. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 57, pp. 2132–2136. SAGE Publications (2013)
53. Schiavo, G., Cappelletti, A., Zancanaro, M.: Engagement recognition using easily detectable behavioral cues. Intell. Artif. 8(2), 197–210 (2014)
54. Riemer, V., Schrader, C.: Impacts of behavioral engagement and self-monitoring on the development of mental models through serious games: inferences from in-game measures. Comput. Hum. Behav. 64, 264–273 (2016)
55. McMahan, T., Parberry, I., Parsons, T.D.: Evaluating player task engagement and arousal using electroencephalography. Procedia Manuf. 3, 2303–2310 (2015)
56. McMahan, T., Parberry, I., Parsons, T.D.: Evaluating electroencephalography engagement indices during video game play. In: Proceedings of the Foundations of Digital Games Conference (2015)
57. Korn, O., Blatz, M., Rees, A., Schaal, J., Schwind, V., Görlich, D.: Procedural content generation for game props? A study on the effects on user experience. Comput. Entertain. (CIE) 15(2), 1 (2017)
58. Cochran, W.G., Cox, G.M.: Experimental designs (1950)