# Not Very Effective: Validity Issues of the Effectance in Games Scale

Nick Ballou
n.b.ballou@qmul.ac.uk
Queen Mary University of London
London, UK

Heiko Breitsohl
heiko.breitsohl@aau.at
University of Klagenfurt
Austria

Dominic Kao
kaod@purdue.edu
Purdue University
USA

Kathrin Gerling
kathrin.gerling@kuleuven.be
KU Leuven
Belgium

Sebastian Deterding
sebastian.deterding@york.ac.uk
University of York
UK

## ABSTRACT

Effectance—the basic positive experience of causing effects—provides a promising explanation for the enjoyment derived from novel low-challenge game genres featuring ample 'juicy' feedback. To date, game researchers have studied effectance using a little-validated 11-item scale developed by Klimmt, Hartmann, and Frey. To test its dimensionality and discriminant validity, we conducted an online survey (n = 467) asking people to report on effectance and related experiences in a recent play session. Confirmatory and exploratory factor analyses show poor fit with a unidimensional factor structure and poor discriminant validity with common enjoyment and mastery/competence measures, likely due to reverse-coded items and a separable input lag factor. We discuss further possible validity issues like questionable content validity, advise against using the scale in its present form, and close with recommendations for future scale development and use.

## CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; *HCI theory, concepts and models.*

## KEYWORDS

effectance; games; player experience; measurement validity

**ACM Reference Format:**
Nick Ballou, Heiko Breitsohl, Dominic Kao, Kathrin Gerling, and Sebastian Deterding. 2021. Not Very Effective: Validity Issues of the Effectance in Games Scale. In *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '21), October 18–21, 2021, Virtual Event, Austria.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3450337.3483492

## 1 INTRODUCTION

A truism of games research is that in gameplay, players tackle non-trivial challenges for the experience of competence and attentive absorption. This truism is challenged by games in highly popular emerging genres like idle and hyper-casual games such as *Clicker Heroes*, in which players repetitively perform trivial tasks like clicking a stationary figure [1, 7]. These offer little immediately discernible challenge or difficulty curve, which standard theories (like flow of self-determination theory) would predict are required for optimally engaging and enjoyable gameplay [5, 42]. While players can and do seek self-devised challenges in casual and idle games [1, 16], we also know that many players of these games do not [7]. This raises the question what makes such 'low-challenge' games enjoyable and engaging.

One possible answer is *effectance*, a supposed basic positive experience and motivation to cause change in the world [17, 46]: exploding barrels with a single click or triggering fireworks of jewels and coins with a swipe make for a positive player experience, regardless of whether this demonstrates ability or demands attention. Casual and hyper-casual games are known to feature ample *juicy feedback*: "excessive positive feedback" [16, p. 45] on every input that arguably amplifies such experiences of causing effects. While researchers have recently offered more differentiated models of juiciness as an aspect of game feel [13], they still recognise exaggeration of reactions to actions as a key factor of juiciness. Thus, effectance provides one plausible candidate explanation for the appeal of idle and hyper-casual games that are low in challenge, but rich in 'juicy' feedback.

Experimentally testing this explanation requires a way of measuring effectance at a situational, moment-to-moment state level. To our knowledge, the only such measurement is the effectance scale developed by Klimmt, Hartmann, and Frey for video game play, first published in 2007, which features 11 self-report items like "the game responded immediately to my inputs" (see Table 1) [19].

Games researchers have since used the scale to assess self-efficacy in games [41], player agency when engaging with game AI [8], interactivity in VR games [38], and have adapted it to measure effectance experiences of live-streaming audiences when directly or indirectly impacting on-stream content [4, 45]. The scale has found particular application among games researchers studying interactive digital narratives (IDN), where Klimmt and colleagues [21] posited effectance as a key mediator between antecedent features (like usability and believability) and consequent experiences (like curiosity, suspense, or surprise; see also [31]). Based on this, Vermeulen and colleagues [44] developed a measurement battery for the user experience of IDNs, which includes effectance as a


*CHI PLAY '21, October 18–21, 2021, Virtual Event, Austria*

ACM ISBN 978-1-4503-8356-1/21/10.
https://doi.org/10.1145/3450337.3483492

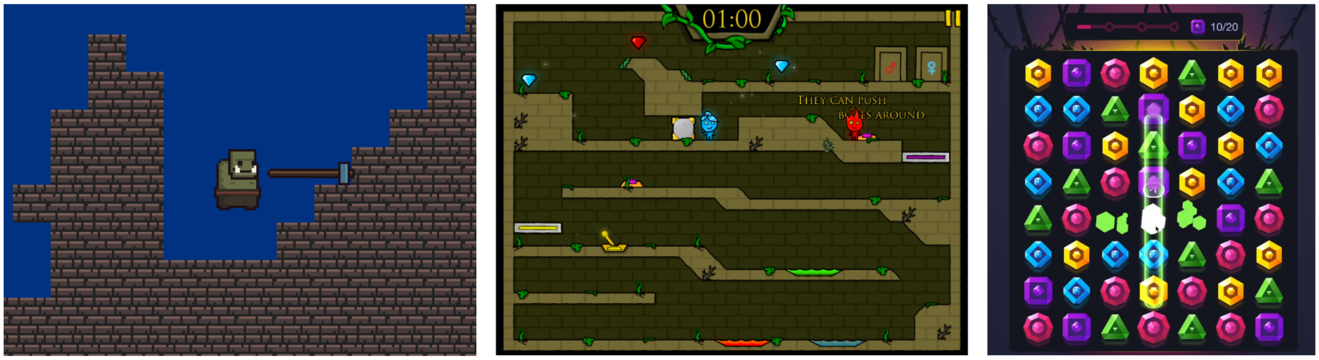


**Figure 1: Screenshots of the three games in the gameplay condition. Left: Getting Ogre it, Center: Fireboy and Watergirl, Right: Jewels FRVR.**

construct, measured by the 11-item effectance scale. A number of studies have since used this battery to test the impact of different user roles [33], replay [32], or watching versus playing [30] on IDN experiences.

Despite its sustained use, the 11-item effectance scale has not been substantively validated. Klimmt et al. [19] reported a Cronbach's $\alpha$ = .89 (n = 500), with Vermeulen et al. [44] later reporting $\alpha$ =.89 (n = 80), indicating sufficient internal consistency for basic research [28, p. 264-265]. However, research shows that internal consistency is not sufficient evidence of validity, as (among other reasons) it is positively related to the length of the scale, and high values do not necessarily indicate that the items are unidimensional [37, 39].

We therefore conducted an online survey (n = 467) asking people to report on effectance and related play experiences in a recent play session to assess scale dimensionality and discriminant validity. We found poor fit with a unidimensional factor structure and high correlations with competence and enjoyment measures. Exploratory factor analyses suggest a hard to interpret two-factor structure, even when discounting problematic reverse-coded items (i.e., negatively-worded items). We conclude that we cannot unconditionally recommend using the current 11-item effectance scale and propose a smaller 4-item subset for future scale development and validation work.

## 2 BACKGROUND

The construct of effectance was introduced to games research by Christoph Klimmt [17, 20]. Klimmt posits that entertainment experiences in video game play arise from multiple synergistic psychological processes operating at different temporal and structural scales. At the molecular level of moment-to-moment input-output loops, players can experience enjoyable *effectance*, the basic sensation of our actions causing change in the world: this experience is found most purely in the "joy in being a cause" (Groos in [46, p. 316]) of children's sensorimotor play, like jumping into puddles or popping bubble-wrap.

Video games afford such effectance through their *interactivity*, specifically the temporal contiguity and disproportionality of input/output. That is, feedback is immediate and out-sized relative to the player action [17, p. 76-81] – what researchers and developers have called "juiciness" [13, 16].

Klimmt (and others after him) see effectance as a component of *agency* experiences [20, 31], and directly linked to *self-efficacy*, defined as "people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives" [2, p. 2]. For these scholars, self-efficacy is the learned expectation of personal effectance, built up through repeated actual effectance experiences [18, 41].

As noted, Klimmt distinguishes effectance from control [9] or competence as conceptualised in e.g. self-determination theory [35]. While effectance is concerned with the moment-to-moment joy of causing effects, which might even be unplanned, control or competence arise at the molar, episode-to-episode level of deliberately setting, pursuing, and attaining goals [17, p. 81-95]: observing goal attainment leads players to ascribe capability to themselves, which is positively experienced and satisfies the need for competence. Practically, effectance and competence often go hand in hand.

Klimmt traces his conception back to White's [46] seminal paper that introduced "effectance" in psychology. For White, effectance is an intrinsic need driving "competence" as a developmental outcome: in play, humans and animals immediately seek and experience the positive satisfaction of a "feeling of efficacy" [46, p. 322] through cycles of focused exploratory behaviour which fuels learning: the build-up of increasing competence to effectively engage the world. Importantly, where Klimmt conceptualises effectance and competence as distinct motives among many others, White posited effectance as the single umbrella intrinsic motivation, while self-determination theory—itself expressly building on White—subsumes effectance as an aspect of competence [35, p. 11,95].

## 3 METHOD

To facilitate the rigorous measurement of effectance, we therefore conducted a validation study of Klimmt et al.'s 11-item scale [19]. Our study received ethical approval from Purdue University (IRB #1809020998). The data, analysis code, and study materials can be found at https://osf.io/6gmw2/.

**Table 1: Items in the re-translated Effectance in Games Scale**

| Item | Wording |
|---|---|
| 1 | The game responded immediately to my inputs. |
| 2 | My inputs had little impact on what happened in the game. (R) |
| 3 | It seemed to me that the game does whatever it wants. (R) |
| 4 | The game responded sluggishly to my inputs. (R) |
| **5** | **What happened in the game was controlled by me.** |
| 6 | Sometimes I couldn't tell whether an event was triggered by me, or whether something else caused it. (R) |
| **7** | **I had the feeling that my inputs directly impacted what was displayed on the screen.** |
| 8 | In part, my inputs had no noticeable impact on what happened in the game. (R) |
| **9** | **It was clear which events I triggered through my inputs.** |
| 10 | My inputs were executed in the game without delay. |
| **11** | **The results of my inputs were clearly recognisable.** |

*(R) indicates reverse-scored items. Bolded items (5, 7, 9, 11) are proposed for future exploration on measuring effectance experiences after gameplay.*

## 3.1 Procedure

We chose a sample size to provide 90% power to detect measurement invariance of $\Delta$RMSEA $\geq$ .007 [26], which yielded a minimum sample of 470. We recruited participants using Prolific.co in June 2021, screening for people who list video games as an activity they engage in with no restriction on self-reported time spent gaming. Participants were randomly assigned to one of two conditions: gameplay and recall. While recalled recent play experiences is a common data source for scale validation in games research (e.g.[6]), as they provide a great and ecological valid variety of play experiences, they are also less controlled and face common memory biases. We therefore chose to combine recall data with data from players who were instructed to play a game just before answering the survey.

In this latter gameplay condition, participants were randomized to play one of three browser-based games for 4 minutes (Figure 1): *Getting Ogre It*, a riff on the game *Getting Over It with Bennett Foddy* featuring intentionally difficult controls for maneuvering a character across obstacles,[1] *Fireboy and Watergirl*, a simple 2D puzzle platformer,[2] and *Jewels FRVR*, a very "juicy" Match 3 puzzle game.[3] We chose these three games to create variance in the data that would plausibly discriminate between the different constructs we compared: *Jewels FRVR* should be high on effectance but not mastery (as it features little challenge, but a large amount audiovisual feedback such as tiles exploding and encouragement from the announcer when a match is made). *Fireboy and Watergirl* should be higher on mastery and curiosity (featuring interesting and challenging puzzles), but lower on effectance as it features less juicy and frequent feedback on actuating game controls. *Getting Ogre It* we expected to be lower in effectance and mastery (due to its difficult controls) but higher in curiosity (due to novel gameplay and interesting obstacles).

In the recall condition, participants were given the prompt "Please take a moment to think carefully about the most recent time you played a video game. What happened in the game? How did you feel while playing it?" In either condition, after playing or seeing the prompt, participants proceeded to the questionnaire, and were instructed to complete the items with reference to the game they just played or the game experience they had recalled, respectively.

[1]https://jul2040.itch.io/gettingogreit
[2]https://www.coolmathgames.com/0-fireboy-and-water-girl-in-the-forest-temple
[3]https://games.crazygames.com/en_US/jewels-frvr/index.html

## 3.2 Measures

Effectance was measured with the 11-item scale developed by Klimmt et al. [19], which its authors provided us in German (the language of their study) and an English translation. As we found some issues with the provided translation, we first re-translated the German items into English using the Translate, Review, Adjudicate, Pretest, and Document (TRAPD) procedure [10, 11]. Two researchers, both native German speakers and native-level English speakers, independently translated the items into English. Two separate researchers, one native German-speaking and one native English-speaking, then reviewed the two candidate translations and adjudicated between them.

We then ran a cognitive pretest with the resultant items. Three participants, each having just played one of the aforementioned three games, completed the scale while thinking aloud and were then asked how intelligible, comprehensive, and long they found the scale. Participants did not raise any significant issues with comprehension, but did flag minor concerns about redundant items. The final wording of each item can be found in Table 1.

In addition to effectance, we measured three constructs that following White [46] should be part of effectance, and following Klimmt [20] should be distinct from effectance, which makes them

**Table 2: Descriptive Statistics**

| | Gameplay (n = 237) | | | Recall (n = 230) | | |
|---|---|---|---|---|---|---|
| | Median | Mean | SD | Median | Mean | SD |
| Age | 22.00 | 24.00 | 6.03 | 23.00 | 24.26 | 5.59 |
| Effectance | 5.27 | 5.14 | 1.20 | 5.73 | 5.65 | 0.90 |
| Mastery | 4.67 | 4.32 | 1.65 | 5.67 | 5.44 | 1.15 |
| Curiosity | 5.33 | 4.88 | 1.65 | 5.67 | 5.36 | 1.40 |
| Enjoyment | 5.29 | 4.94 | 1.62 | 5.86 | 5.79 | 0.94 |

**Table 3: Item Loadings in the Exploratory Factor Analysis**

| | Gameplay | | | Recall | | | Combined | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Factor 1** | **Factor 2** | **Uniq.** | **Factor 1** | **Factor 2** | **Uniq.** | **Factor 1** | **Factor 2** | **Uniq.** |
| eff1 | | 0.967 | 0.170 | 0.814 | | 0.372 | 1.030 | | 0.169 |
| eff2r | 0.679 | | 0.583 | | 0.695 | 0.558 | | 0.675 | 0.593 |
| eff3r | 0.743 | | 0.528 | | 0.492 | 0.745 | | 0.690 | 0.598 |
| eff4r | | 0.663 | 0.485 | | 0.603 | 0.518 | 0.550 | | 0.557 |
| eff5 | 0.805 | | 0.322 | | | 0.852 | | 0.525 | 0.655 |
| eff6r | 0.794 | | 0.498 | | | 0.743 | | 0.693 | 0.604 |
| eff7 | 0.535 | | 0.488 | 0.721 | | 0.534 | 0.424 | 0.346 | 0.516 |
| eff8r | 0.481 | | 0.768 | | 0.807 | 0.471 | | 0.548 | 0.712 |
| eff9 | 0.583 | | 0.469 | 0.698 | | 0.516 | | 0.428 | 0.502 |
| eff10 | | 0.953 | 0.243 | 0.617 | | 0.590 | 0.930 | | 0.337 |
| eff11 | 0.534 | | 0.429 | 0.683 | | 0.481 | 0.463 | 0.365 | 0.440 |

*Applied rotation method is promax. Uniq. = uniqueness. Labels in the first column correspond with item numbers in Table 1.*

suitable for testing discriminant validity: mastery, curiosity, and enjoyment. We operationalized mastery and curiosity using their respective subscales from the Player Experience Inventory (PXI) [43], and enjoyment with the interest/enjoyment subscale of the Intrinsic Motivation Inventory (IMI) [34]. The PXI is a notably well-validated and comprehensive player experience instrument [43]. Specifically, its "mastery" subscale shows a strong correlation with the Player Experience of Need Satisfaction [PENS; 36] "competence" subscale, arguably the most frequently used scale for measuring game-related competence [27], but avoids documented issues with the factor structure of PENS [15]. We used the interest/enjoyment IMI subscale as it is well-validated [e.g., 25] and presently among the most frequently direct enjoyment measures in games research [27].

While all three scales originally use a 7-pt Likert scale, the PXI ranges from -3 ("strongly disagree") to 3 ("strongly agree"), and the effectance measure did not specify anchors. To ensure a standard response format across scales, we used the anchors from the IMI. Thus, all items are measured on 7-pt Likert scale from 1 ("not at all true") to 7 ("completely true").

## 3.3 Participants

A total of 470 participants completed the survey. Three participants were removed due to technical problems with the embedded game. We identified an additional 13 participants as possibly untrustworthy because they selected the same response 10 times or more in a row. However, when we manually inspected their responses, we found that the majority responded attentively to reverse-coded items. Also, including or excluding these participants did not meaningfully change the results. Therefore, we report analyses based on all 467 completed surveys. On average, the survey took 8.6 minutes for the gameplay group and 4.5 minutes for the recall group to complete. Participants in the gameplay condition were paid £1.10, and those in the recall condition £0.55. Descriptive statistics for the measures are shown in Table 2.

# 4 RESULTS

Given that there was a clear intended one-factor structure, we began by calculating reliability among the 11 effectance items, which was high (McDonald's $\omega_h = .85, 95\% CI[.81, .88]$, Cronbach's $\alpha = .87$). We then proceeded to conduct a confirmatory factor analysis (CFA) of the 11 effectance items. We intentionally ran a range of common fit indices, namely $\chi^2$, CFI, RMSEA, and SRMR.

While we did not pre-specify cut-off values for fit indices, the CFA results indicated poor fit to the data ($\chi^2$ [44 df] = 394.008, $p < .001$, CFI = .763, RMSEA = .131 90% CI [.121, .141], SRMR = .092): all four indices were considerably worse than commonly-used heuristics for adequate fit [14]. This indicates that a one-factor model does not fit the data.

A common source of model misfit are reverse-coded items, which often correlate with each other and thus introduce multidimensionality [12]. To explore this possibility, we fitted a CFA model with an added factor accounting for covariation among the 5 reverse-scored items. Model fit was significantly improved, but still remained overall poor ($\chi^2$ [39 df] = 288.406, $p < .001$, CFI = .831, RMSEA = .118 90% CI [.107, .129], SRMR = .073).

To understand possible reasons for the poor fit, we followed up with an exploratory factor analysis using promax. Parallel analysis selected a two-factor solution (Table 3), with 3 items loading solely onto the first factor, 6 items loading solely onto a second factor, and two cross-loading items (items that loaded onto both factors ≥ .32, following [40]).

The two factors that emerged were not easily interpretable. We therefore decided to separate the recall and gameplay conditions and run an EFA on each to assess differences between them. Analysis selected two-factor solutions for each of the subgroups. In the gameplay group, three items loaded onto a separate factor one could interpret as input lag, as all of them refer to the speed with which a game responds to player inputs (items 1, 4, 10). In the recall group, the second factor consists of all but one of the reverse-coded items (2, 3, 4, 8).

To test for discriminant validity, we fitted separate CFA models for the gameplay and recall groups, each including the items from all four constructs (all 11 effectance items, curiosity, mastery, and enjoyment, as well as the reverse-coding method factor). Rönkkö and Cho [29] propose that two latent factors can be accepted as sufficiently distinct when latent factor correlations have a 95% confidence interval upper bound that does not exceed .8. Results in the recall group supported the discriminant validity of effectance (curiosity: $r = .19$, 95% CI [.05, .34], mastery: $r = .49$[.34, .64], and enjoyment: $r = .38$[.22, .55]). Results for the gameplay group slightly exceeded this bound, with effectance correlating strongly with both mastery ($r = .74$, 95% CI [.66, .82]) and enjoyment ($r = .72$, 95% CI [.64, .80]), and moderately with curiosity ($r = .53$, 95% CI [.41, .65]). Put differently, the tested operationalization of effectance is not sufficiently distinct from mastery and enjoyment in the gameplay group.

## 5 DISCUSSION AND LIMITATIONS

In sum, our results suggest that the effectance in games scale [19] is unsuitable for use in its current form. Although reliability was acceptable, factor analysis results did not support a one-factor model, and discriminant validity in relation to mastery and enjoyment was poor among players who had just played a game. In other words, the effectance scale may measure more than one separable aspect of player experience, and also cannot be adequately differentiated from other related, but theoretically distinct, constructs. Consequently, previous research using this measure should be interpreted with caution [4, 8, 19, 30, 32, 33, 38, 41, 44, 45].

This finding reinforces the limitations of scale evaluations relying solely on reliability measures like Cronbach's alpha [39]. We found two major sources of multidimensionality that limited the measure's validity. First, reverse-coded items were a major source of misfit that formed their own factor in the recall group: these issues of reverse-coded items are well-known in psychometrics [12] but rarely studied in games research.

Second, we found a separable "input lag" factor in the gameplay group, which pertains to the delay between making a command and the game reacting to that command; this may have manifested as participants played in-browser versions of games that can be sluggish (and *Getting Ogre It* is intentionally cumbersome and slow to control). Input lag is an important issue [22], but to date has received limited research attention [24]. While this may be a worthwhile topic for future work—and could potentially be measured by items 1, 4, and 10 in the effectance scale here—input lag is conceptually distinct from effectance, and on the basis of our data, should not be included in the same construct.

Discriminant validity presents a somewhat more minor issue of the present scale—as applied to a just-played gaming experienced. Theoretically and on face validity, the tested effectance items are clearly separable from competence and enjoyment. Our data showed that these were indeed statistically distinct for the gameplay condition, but only somewhat statistically distinct in the recall condition. We note that scores for all three constructs are heavily skewed towards the upper end of the scales, especially in the recall group; it therefore may be the case that ceiling effects inflated the correlations between factors.

The differences between conditions with regard to both factor analysis results and discriminant validity also highlight the fact that players' responses to a particular state-level construct may be affected by the recency of that state. Participants less clearly discriminated between related but distinct constructs when recalling a prior gameplay experience. These results indicate that the relatively common practice of assessing state-like gameplay experiences in guided recall may at best not be psychometrically comparable with just-lived experiences, and at worst may be biased in systematic ways.

One important limitation is that our scale validation is still incomplete, missing major evaluations recommended in best practices, such as test-retest reliability and criterion validity [3]. Specifically, we have doubts about the scale's content validity—that is, whether it actually captures the construct in question. If effectance is the *inherently enjoyable* experience of causing *effects*, it is noticeable that none of the 11 scale items speak to any positive experience, and only two speak to the kind and size of effect caused (2, 7). On face value, the scale seems to conceptualise the responsiveness and legibility of a game's interface—in other words, antecedent system features that are necessary but on their own likely insufficient preconditions for effectance. Future work should consider expert and target population evaluations of content validity [3].

Should researchers have no other viable alternative and be willing to accept the limitations outlined in this paper, using a subset of the effectance scale we tested may be tenable. In particular, for immediate post-play surveys, we can see using items 5, 7, 9, and 11, which would avoid both reverse-coded items and the "input lag" factor. Again, we caution that without further tests of content and criterion validity, it is unclear whether this item set captures effectance experiences as conceptualised by White or Klimmt, or antecedent features of interface responsiveness and legibility.

We concur with Klimmt and other researchers that effectance is a promising construct for understanding molecular, moment-to-moment user experiences – not just for games featuring juicy feedback, but also as part of the interaction aesthetics [23] of interactive systems more generally. Yet given the issues our limited validation study already found, this promise might be best served by developing and validating a new scale.

## CONFLICTS OF INTEREST

None to declare.

## DATA AVAILABILITY

All data, materials, and code associated with this project are available on the Open Science Framework (https://osf.io/6gmw2/).

## ACKNOWLEDGMENTS

This work was supported by the EPSRC Centre for Doctoral Training in Intelligent Games and Game Intelligence (IGGI) [EP/S022325/1] and the Digital Creativity Labs, funded by EPSRC/AHRC/Innovate UK [EP/M023265/1]. We thank Christoph Klimmt for providing access to the original scale and permission to publish it.

## REFERENCES

[1] Sultan A. Alharthi, Olaa Alsaedi, Zachary O. Toups, Theresa Jean Tanenbaum, and Jessica Hammer. 2018. Playing to Wait: A Taxonomy of Idle Games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–15. https://doi.org/10.1145/3173574.3174195

[2] Albert Bandura. 1994. Self-Efficacy. In *Encyclopedia of Human Behavior*, V S Ramachaudran (Ed.). Vol. 4. Academic Press, New York, 71–81. https://www.uky.edu/~eushe2/Bandura/Bandura1994EHB.pdf

[3] Godfred O Boateng, Torsten B Neilands, Edward A Frongillo, Hugo R Melgar-Quiñonez, and Sera L Young. 2018. Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in public health* 6 (2018), 149.

[4] Simon Bründl, Christian Matt, and Thomas Hess. 2017. Consumer Use of Social Live Streaming Services: The Influence of Co-Experience and Effectance on Enjoyment. In *Proceedings of the 25th European Conference on Information Systems (ECIS'17)*. AIS Electronic Library, Guimarães, 1775–1791.

[5] Jenova Chen. 2007. Flow in games (and everything else). *Commun. ACM* 50, 4 (2007), 31–34.

[6] Alena Denisova, Paul Cairns, Christian Guckelsberger, and David Zendle. 2020. Measuring perceived challenge in digital games: Development & validation of the challenge originating from recent gameplay interaction scale (CORGIS). *International Journal of Human-Computer Studies* 137 (May 2020), 102383. https://doi.org/10.1016/j.ijhcs.2019.102383

[7] facebook gaming. 2020. *Genre and Great Games: Understanding Audiences and Designing Better Mobile Games.* Technical Report. facebook. 88 pages. https://www.facebook.com/fbgaminghome/marketers/genre2020

[8] Francesca Foffano and David Thue. 2019. Changes of User Experience in an Adaptive Game: A Study of an AI Manager. In *FDG '19: Proceedings of the 14th International Conference on the Foundations of Digital Games.* ACM, San Luis Obispo, Art. 99.

[9] Torben Grodal. 2000. Video Games and the Pleasures of Control. In *Media entertainment: The psychology of its appeal*, Dolf Zillmann and Peter Vorderer (Eds.). Lawrence Erlbaum, Mahwah, NJ, 197–213.

[10] Janet Harkness, Beth-Ellen Pennell, and Alisú Schoua-Glusberg. 2004. Survey Questionnaire Translation and Assessment. In *Wiley Series in Survey Methodology*, Stanley Presser, Jennifer M. Rothgeb, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer (Eds.). John Wiley & Sons, Inc., Hoboken, NJ, USA, 453–473. https://doi.org/10.1002/0471654728.ch22

[11] Janet A. Harkness. 2003. Questionnaire Translation. In *Cross-Cultural Survey Methods*, Janet A. Harkness, Fons J. R. van de Vijver, and Peter Ph Mohler (Eds.). J. Wiley, Hoboken, N.J, 35–56.

[12] J. Herche and B. Engelland. 1996. Reversed-Polarity Items and Scale Unidimensionality. *Journal of the Academy of Marketing Science* 24, 4 (Sept. 1996), 366–374. https://doi.org/10.1177/0092070396244007

[13] Kieran Hicks, Patrick Dickinson, Juicy Holopainen, and Kathrin Gerling. 2018. Good Game Feel: An Empirically Grounded Framework for Juicy Design. In *Proceedings of DiGRA 2018.* Digital Games Research Association, Turin, 1–17.

[14] Li-tze Hu and Peter M. Bentler. 1999. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 6, 1 (Jan. 1999), 1–55. https://doi.org/10.1080/10705519909540118

[15] Daniel Johnson, M John Gardner, and Ryan Perry. 2018. Validation of Two Game Experience Scales: The Player Experience of Need Satisfaction (PENS) and Game Experience Questionnaire (GEQ). *International Journal of Human-Computer Studies* 118 (Oct. 2018), 38–46. https://doi.org/10.1016/j.ijhcs.2018.05.003

[16] Jesper Juul. 2011. *A Casual Revolution: Reinventing Video Games and Their Players.* MIT Press, Cambridge, MA; London.

[17] Christoph Klimmt. 2006. *Computerspielen als Handeln: Dimensionen und Determinanten des Erlebens interaktiver Unterhaltungsangebote.* Herbert von Halem, Köln.

[18] Christoph Klimmt and Tilo Hartmann. 2006. Effectance, Self-Efficacy, and the Motivation to Play Video Games. In *Playing Video Games: Motives, Responses, and Consequences*, Peter Vorderer and Jennings Bryant (Eds.). Lawrence Erlbaum, Mahwah, NJ; London, 133–145.

[19] Christoph Klimmt, Tilo Hartmann, and Andreas Frey. 2007. Effectance and Control as Determinants of Video Game Enjoyment. *CyberPsychology & Behavior* 10, 6 (Dec. 2007), 845–848. https://doi.org/10.1089/cpb.2007.9942

[20] Christoph Klimmt and Daniel Possler. 2021. A Synergistic Multiprocess Model of Video Game Entertainment. In *The Oxford Handbook of Entertainment Theory*, Peter Vorderer and Christoph Klimmt (Eds.). Oxford University Press, Oxford, 623–646.

[21] Christoph Klimmt, Christian Roth, Ivar Vermeulen, Peter Vorderer, and Franziska Susanne Roth. 2012. Forecasting the Experience of Future Entertainment Technology. *Games and Culture* 7, 3 (2012), 187–208.

[22] Richard Leadbetter. 2009. Console Gaming: The Lag Factor. https://www.eurogamer.net/articles/digitalfoundry-lag-factor-article

[23] Eva Lenz, Sarah Diefenbach, and Marc Hassenzahl. 2014. Aesthetics of interaction: a literature synthesis. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational.* ACM, Helsinki, 628–637.

[24] Chris Lewis, Jim Whitehead, and Noah Wardrip-Fruin. 2010. What Went Wrong: A Taxonomy of Video Game Bugs. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games - FDG '10.* ACM Press, Monterey, California, 108–115. https://doi.org/10.1145/1822348.1822363

[25] Edward McAuley, Terry Duncan, and Vance V. Tammen. 1989. Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport* 60, 1 (March 1989), 48–58. https://doi.org/10.1080/02701367.1989.10607413

[26] Adam W. Meade, Emily C. Johnson, and Phillip W. Braddy. 2008. Power and Sensitivity of Alternative Fit Indices in Tests of Measurement Invariance. *Journal of Applied Psychology* 93, 3 (May 2008), 568–592. https://doi.org/10.1037/0021-9010.93.3.568

[27] Elisa D Mekler, Julia Ayumi Bopp, Alexandre N Tuch, and Klaus Opwis. 2014. A Systematic Review of Quantitative Studies on the Enjoyment of Digital Entertainment Games. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems.* ACM Press, Toronto, 927–936. https://doi.org/10.1145/2556288.2557078

[28] Jum C Nunnally and Ira H Bernstein. 1994. *Psychometric theory* (3rd ed.). McGraw-Hill, New York. OCLC: 28221417.

[29] Mikko Rönkkö and Eunseong Cho. 2020. An Updated Guideline for Assessing Discriminant Validity. *Organizational Research Methods* 1, 1 (Nov. 2020), 1–42. https://doi.org/10.1177/1094428120968614

[30] Christian Roth, Christoph Klimmt, Ivar E Vermeulen, and Peter Vorderer. 2011. The Experience of Interactive Storytelling: Comparing "Fahrenheit" with "Façade". In *10th International Conference on Entertainment Computing, ICEC 2011.* Springer, Vancouver, BC, 13–21.

[31] Christian Roth and Hartmut Koenitz. 2016. Evaluating the User Experience of Interactive Digital Narrative. In *AltMM'16: Proceedings of the 1st International Workshop on Multimedia Alternate Realities.* ACM, Amsterdam, 31–36.

[32] Christian Roth, Ivar Vermeulen, Peter Vorderer, and Christoph Klimmt. 2012. Exploring Replay Value: Shifts and Continuities in User Experiences Between First and Second Exposure to an Interactive Story. *CyberPsychology, Behavior & Social Networking* 15, 7 (2012), 378–381.

[33] Christian Roth, Ivar Vermeulen, Peter Vorderer, Christoph Klimmt, David Pizzi, Jean-Luc Lugrin, and Marc Cavazza. 2012. Playing In or Out of Character: User Role Differences in the Experience of Interactive Storytelling. *CyberPsychology, Behavior & Social Networking* 15, 11 (2012), 630–633.

[34] Richard M. Ryan. 1982. Control and Information in the Intrapersonal Sphere: An Extension of Cognitive Evaluation Theory. *Journal of Personality and Social Psychology* 43, 3 (1982), 450–461. https://doi.org/10.1037/0022-3514.43.3.450

[35] Richard M Ryan and Edward L Deci. 2017. *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness.* Guilford Press, New York.

[36] 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, 4 (Dec. 2006), 344–360. https://doi.org/10.1007/s11031-006-9051-8

[37] Neal Schmitt. 1996. Uses and Abuses of Coefficient Alpha. *Psychological Assessment* 8, 4 (1996), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

[38] Daniel M Shafer, Corey P Carbonara, and Michael F Korpi. 2019. Factors Affecting Enjoyment of Virtual Reality Games: A Comparison Involving Consumer-Grade Virtual Reality Technology. *Games for Health Journal* 8, 1 (2019), 15–23.

[39] Leanne M. Stanley and Michael C. Edwards. 2016. Reliability and Model Fit. *Educational and Psychological Measurement* 76, 6 (Dec. 2016), 976–985. https://doi.org/10.1177/0013164416638900

[40] Barbara G. Tabachnick and Linda S. Fidell. 2001. *Using Multivariate Statistics* (4th ed ed.). Allyn and Bacon, Boston, MA.

[41] Sabine Trepte and Leonard Reinecke. 2011. The Pleasures of Success: Game-Related Efficacy Experiences as a Mediator Between Player Performance and Game Enjoyment. *CyberPsychology, Behavior & Social Networking* 14, 9 (2011), 555–557.

[42] April Tyack and Elisa D Mekler. 2020. Self-determination theory in HCI games research: current uses and open questions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* ACM, Honolulu, 1–22. https://doi.org/10.1145/3313831.3376723

[43] Vero Vanden Abeele, Katta Spiel, Lennart Nacke, Daniel Johnson, and Kathrin Gerling. 2020. Development and Validation of the Player Experience Inventory: A Scale to Measure Player Experiences at the Level of Functional and Psychosocial Consequences. *International Journal of Human-Computer Studies* 135 (March 2020), 102370. https://doi.org/10.1016/j.ijhcs.2019.102370

[44] Ivar E Vermeulen, Christian Roth, Peter Vorderer, and Christoph Klimmt. 2010. Measuring User Responses to Interactive Stories: Towards a Standardized Assessment Tool. In *Third Joint Conference on Interactive Digital Storytelling, ICIDS 2010.* Springer, Berlin, Heidelberg, 38–43.

[45] Kun Wang, Zhao Pan, Yaobin Lu, and Sumeet Gupta. 2020. What Motives Users to Participate in Danmu on Live Streaming Platforms? The Impact of Technical Environment and Effectance. *Data and Information Management* 3, 3 (2020), 117–134.

[46] R W White. 1959. Motivation reconsidered: the concept of competence. *Psychological review* 66 (Sept. 1959), 297–333. http://www.ncbi.nlm.nih.gov/pubmed/13844397