

Get It Right! Introducing a Framework for Integrating Validation in Applied Game Design

Harald Warmelink^(✉), Marilla Valente, Richard van Tol,
and Robbertjan Schravenhoff

HKU Innovation Studio, HKU University of the Arts Utrecht,
Lange Viestraat 2b, 3511 BK Utrecht, The Netherlands
{harald.warmelink, marilla.valente, richard.vantol,
robbertjan.schravenhoff}@hku.nl

Abstract. Validation of applied games is generally a lengthy and costly process, usually done after the game has (almost) been completed. This article argues that validation throughout the design process is vital for creating games that achieve their desired objective. It proposes a framework to integrate validation throughout the applied game design process. Iterative design and design fidelity levels form the foundation of the framework. In the framework four types of design fidelity (specifications, concepts, prototypes, and integrations) are linked to five types of validity (content, face, construct, concurrent and predictive validity), which lead to eight applicable validation tools and techniques. The framework is a starting point for further research into applied game validation.

1 Introduction

Recently many have argued that validation should be a prime concern for current serious, simulation, educational and applied game designers and researchers [1–3]. Designers need to know if and how their games achieve their objectives in order to act responsibly. Researchers need to know the same in order to advance game design theory and practice [3, 4]. Over the past decade the research community has already proven to be invaluable in this respect by providing over 11 different frameworks for game evaluation and assessment [3, 5, 6].

While the work on game evaluation and assessment is an important step, they have their limits as validation strategies. First, the late timing of this kind of validation increases the chance of disappointment. Full-scale evaluation or assessment in accordance to e.g. Mayer et al. [3] is only possible after a costly game design and development project has all but finished. Second, evaluation or assessment connotes a problematic division between design and validation, or designer and researcher. Third, evaluation or assessment tends to downplay the multifaceted and intricate nature of game design and thus validation. Game design involves a high amount of intertwined variables (e.g. audiovisual, artistic or narrative design, the design of goals, rules and feedback systems). In some frameworks many of these variables are problematically

grouped together (into e.g. ‘game/simulation quality’; [6]) and are at best ‘tweakable’ variables by the time of an evaluation or assessment.

This article therefore extends existing work by exploring how validation can be integrated into the game design process. It offers a framework for integrating validation during game design as a starting point for further research both into the framework’s foundations and into potential tools and techniques of validation. We do so not only to contribute to the continually developing field of game design research, but also to aid game design education.

Throughout this article we consistently use the term applied games as we consider it a useful umbrella term, covering games as digital, analog or hybrid interactive systems of trial and error that players engage with in order to learn or be enticed into certain behavior. In doing so we acknowledge the valuable work of the different communities of serious, simulation, educational and applied game design research.

Our understanding of validation follows Graafland et al., defining validity as whether “an instructional instrument (such as a serious game) adequately resembles the construct it aims to educate or measure” [2]. In doing so we acknowledge particularly the medical, pharmaceutical and simulation design fields as having a long-standing history concerning validation trials, in general, with simulations and with games [2, 7].

Our validation framework is based on a literature review, two years of research-through-design, and reflections on applied game design projects we were involved in. The literature review is based on evaluation and design frameworks from the applied gaming research community, as well as validation frameworks of the aforementioned related disciplines. The two years of research-through-design focused on the creation of tools (e.g. a prototype of a short card game) for applied game designers not (fully) aware of the nature and need of validation. This tool creation process helped us ascertain different timings and levels of validation during an applied game design process.

2 The Framework’s Foundations

2.1 Iterative Design

Our validation framework is based on an *iterative* approach to game design. In iterative design cycles a typical ‘testing’ or ‘check’ phase is or at least should be a form of validation (Fig. 1). After all, put most simply, an iterative design cycle entails conceptualising a game experience, building it, and testing what has been built in order to ascertain whether it accomplishes what was conceptualised.

Although an iterative design approach is arguably commonplace, its nature and value is too easily overlooked. We conceptualise applied game designers as aiming for novel, creative designs that motivate certain target audiences into learning or specific behavior. This means that we consider it possible to find a design that meets self-defined criteria related not only to the main objective at hand, but to state-of-the-art game design and to what motivates the target audience as well. An iterative design approach is key to reaching these particular, multiple objectives [8–10].

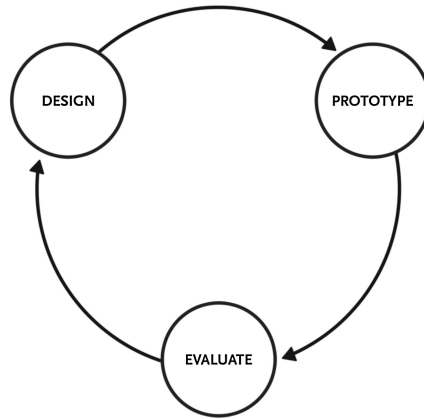


Fig. 1. Fundamental phases of the iterative design cycle, forming our framework’s foundation.

Based on the iterative design cycle, our framework distinguishes three consecutive, reoccurring phases in the applied game design process: design, build and validate. The first phase, *design*, is one where specifications, concepts or prototypes are thought up, often through creativity-boosting means (e.g. the use of an ideation technique such as group brainstorming). The second phase, *build*, is one where these specifications, concepts or prototypes are developed into artefacts such as design documents and paper prototypes. The third phase, *validate*, is the phase in which these artefacts are actually validated - the focal point of this article.

2.2 Design Fidelity

Our framework also incorporates a second dimension, one depicting the gradual move towards the required *design fidelity*. With design fidelity we refer to the level of proximity a design has to the envisioned final product. For example, a paper prototype has higher design fidelity than a graphical user-interface sketch or wireframe, while such a sketch or wireframe has higher design fidelity than a design requirements list.

The inclusion of the design fidelity dimension has important consequences for our understanding of the design process, and the role of validation in it. By combining design phases with levels of design fidelity the framework accentuates the validation of multiple and different artefacts during the design process. As such the design fidelity dimension diminishes the gap between the fields of social science and design practice. It opens up new focal points for validation that have previously been overlooked in applied game validation research.

The framework depicts four levels of design fidelity: *Specifications* level, *Concepts* level, *Prototypes* level and *Integrations* level (Fig. 2). *Specifications* denotes the basic specifications or requirements the final design should meet. *Concepts* denotes the grand, basic ideas for the applied game experience, often using common, current game heuristics, e.g., well-known game genre typologies (‘real-time strategy’, ‘simulation’, ‘role-playing’ games). The *Prototypes* level encompasses several design fidelity levels.

For example, when designing a computer game, a paper prototype is of lower design fidelity than a digital prototype. The fourth and final level of *Integrations* depicted in Fig. 2 is the highest level of design fidelity. Once an applied game or any product is fully integrated into a context of use (e.g., an educational setting, or an organisational process), it can be considered a finished product. Hence the center of Fig. 2 is also marked with ‘*Finish*’.

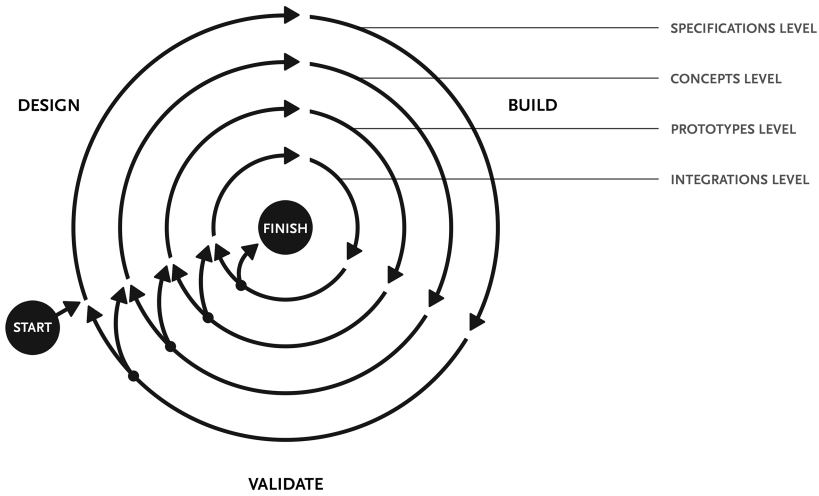


Fig. 2. Integrating validation in applied game design based on four design fidelity levels.

3 The Framework Completed

3.1 Validity Types

Because of the existence of different levels of design fidelity, we propose that applied game design validity has many faces and can take place throughout the entire applied game design process. Graafland et al. [2] provided the inspiration for what validation specifically can entail. Based on the guidelines set forth by The American Psychological Association [11], and mirroring simulation validity types [7], they state that validity research in medical education usually involves several phases and includes *content*, *face*, *construct*, *concurrent*, and *predictive* validity:

“First, experts should scrutinize the game’s content to determine its legitimacy (*content* validity). Second, experts and novices judge the instrument’s apparent similarity to the construct it attempts to represent (*face* validity). *Construct* validity reflects the ability of the instrument to actually measure what it intends to measure (i.e., the difference in performance between groups of users with different levels of experience in reality). *Concurrent* validity reflects the correlation between performance on the serious game and their performance on an instrument believed to measure the same construct (e.g., a simulator or course). The ultimate goal is to prove a game’s *predictive* validity: does performance in the game lead to better outcomes in reality?” [2].

3.2 Validation Tools and Techniques

Table 1 integrates design fidelity and the five validity types to propose a set of tools and techniques applied game designers can use during the validation phase of a particular fidelity level in the design process. In the remainder of this section we explain each of the proposed tools and techniques.

Table 1. Validation tools and techniques for each design fidelity level

Design fidelity level (low to high)	Validity type	Validation tool or technique
Specifications	Content validity: Is the game's content legitimate? How is the content validated to be complete, correct, and nothing but for the intended designated construct?	1. Design requirement framework or heuristic check 2. Subject matter check 3. SMART goal definition check
Concepts	Face validity: Is there a similarity to the construct it attempts to measure? Do educators and trainees view it as a valid way of instruction?	4. Specifications check 5. Applied game design framework or heuristics check
Prototypes	Construct validity: The ability of the instrument to actually measure what it intends to measure. Is the game able to measure different skills it intends to measure?	Same as listed under Concepts, plus: 6. Playtest, with at least an appropriate outcome variable measurement
Integrations	Concurrent validity: The correlation between performance on the designed game and performance on an instrument. How does learning outcome compare to other methods assessing the same medical construct?	Same as listed under Prototypes, plus either/both: 7. A (quasi-)experiment or randomised controlled trial
	Predictive validity: Does performance in the game lead to better outcomes in reality? Does playing the game predict skills improvement in real life?	8. Stealth assessment: behavioral data gathering and analysis

Design Requirement Framework or Heuristic Check. To ensure that the design team has defined the right design requirements, we propose that the team juxtaposes their proposed requirements onto the following simple requirement framework/heuristic. Thus the team asks, critically, ‘have we covered all possible requirements in our design requirements list?’

Technological, Institutional and Legal/Policy Constraints. The preferred or required context of use will impose technological, institutional and legal constraints. For example, within the domain of education, a formal educational institute will always

have certain computer hardware (e.g. PCs, smartboards) and facilities (e.g. computer labs) at its disposal that the team will at least need to be familiar with. They might also have a support staff for equipment lending, troubleshooting or security, complete with undefiable protocols and procedures. There might also be relevant regional or national legislation or policies in place.

Target Audience Specifics. An applied game design team designs for one or more specific target audiences who should be engaged by the game, intrinsically and/or extrinsically. This obligates the design team to know all about their target audiences. Without enough insights, the design team runs the risk of designing a product that falls short of the expectations. One common technique for defining for whom the designer actually designs is the development of one or more personas. Here personas are defined as ‘a documented set of archetypal people who are involved with a product or service’ [12].

Subject Matter Check. All applied games have subject matter, regardless of the domain in question (e.g. health, education). The subject matter behind an applied game for education might be medieval history of a certain country, or decision-making processes in national government, for example. Behind an applied game for health might be a tried-and-true physical or mental therapy model. In any case, the subject matter of the applied game design project will be easily identified, yet hard to define concretely. It nevertheless needs to be checked, probably through literature studies with or without the help of a subject matter expert.

SMART Goal Definition Check. Despite different interpretations of the acronym, SMART is essentially a validation technique for goal definitions, allowing for the design team to focus and collaborate. Our interpretation of the often-used SMART acronym is as follows:

- Specific, i.e., the goal is so specific that there is (seemingly) no room for mis- or reinterpretation.
- Measurable or Meaningful, i.e., the goal includes a quantified or qualified level of achievement, either in absolute or relative terms (e.g. playing the applied game should be more effective than some alternative).
- Achievable, i.e., based on their individual expertise, all design team members agree that the goal is actually attainable for the design team in question, given the allotted or preferred time frame.
- Relevant, i.e., the goal entails fulfilling a need or solving a problem through the acquisition of or a change in knowledge, skills, attitude or behavior, and does not encompass the exact nature of the design (e.g. not ‘an online multiplayer computer game’).
- Time-bound, i.e., a (rough) deadline for reaching the goal has been set and can be achieved within the allotted time.

The Achievable and Relevant criteria of the SMART goal validation technique require not only practical design experience, they also require the design team to do further research into disciplines relevant to the goal at hand and/or to collaborate with subject matter experts. For example, if the design team wishes to achieve a long-term behavioral change, the team would need to consult (recent) psychological studies and

theories into the complexities of such an endeavor in order to ascertain whether and how it might actually be achievable within the desired time frame.

Specifications Check. Once the design requirements, subject matter and design goal have been validated, a concept or prototype can in turn be juxtaposed onto them as a next validation step. The design team thus simply asks, critically, ‘does the concept or prototype fit the design requirements, subject matter and design goal that have been set?’

Applied Game Design Framework or Heuristics Check. Several frameworks, heuristics or design principles have been developed and published that applied game designers can use to validate their own concepts or prototypes. Such a validation entails juxtaposing the concept or prototype onto a chosen framework. The domain of the concept or prototype at hand (e.g. health or education) dictates which framework the design team could choose. Possible frameworks include:

- The Four Dimensional Framework [13] or Marne et al.’s framework [14] that forces an applied game designer to make several (e.g. pedagogical) foundations or facets explicit;
- The Triadic Game Design or similar frameworks [15–17], focusing on finding a balance between different requirements emanating from the ‘worlds’ of play, meaning and reality;
- More specific design principles by Dondlinger [18], Charsky [19] or Annetta [20], focusing on e.g. challenges, increased complexity or the use of avatars.

Playtest, with at Least an Appropriate Outcome Variable Measurement. Once a paper or digital prototype has been made, it can be tested. Such a test is essentially a strategy for validating the prototype. Usually this is done by structured observation of the play test (preferably by two observers using the same protocol). Primary concern during a play test is to ascertain to what extent the desired goal of the applied game is reached. The design team needs an indication if this first paper prototype reaches the desired goal for at least the majority of players. If not, it is back to the drawing board.

Yet structured observations of the knowledge, skills, attitude or behavior that the prototype should aim for might not be enough. Most times results of particularly first play tests will be ambivalent. Some players might do what is expected, many might not.

The question that then remains is simply, why? In order to even come close to an answer to this basic yet grand question, more data is required than provided by a structured observation as described above. It needs to be complemented by more qualitative or quantitative data obtained from open or semi-structured interviews or questionnaire. In-game data capturing might also be used. Existing evaluation and assessment frameworks offer ample insights into what data should be obtained [3, 5, 6].

A (Quasi-)Experiment or Randomised controlled trial. If part of the design goal is to integrate a game in a certain context that serves its purpose better than its predecessor, a (quasi-)experiment or randomised controlled trial is the best final validation technique to apply. Think of an educational game that is meant to replace a lecture (series), since it is assumed that playing the game will have the students reach the learning objective more often than the lecture (series).

Often a quasi-experiment is the most practical to set up and also rigorous enough and most relevant for the purpose at hand. The design team forms two groups of comparable size and composition, and has the first group play the game, while the second group serves as the control group, doing the original activity that the game is meant to replace. A third comparable group might also be formed as another control group, doing something else entirely or even nothing at all, if the design team wants to ascertain whether the game not only performs better than the original activity, but that it has any added value to begin with. In any case, the same measurements must be applied to all groups, which includes a baseline measurement prior to the intervention (game or otherwise).

For an applied game where the stakes are extremely high, e.g. in the domain of health [21], a randomised controlled trial is probably essential, although it may require the help of a research specialist. This full-experiment setup has additional rigor thanks to its larger scale (i.e., it should involve many more participants with more diverse backgrounds), more longitudinal focus (multiple post-game/-intervention measurements are carried out over a period of days/weeks/months) and its stricter protocols concerning the assignment of participants (i.e., study personnel is unaware of the actual assignment).

Stealth Assessment: Behavioral Data Gathering and Analysis. To validate a full integration or implementation of an applied game, one can gather and analyze behavioral data obtained ‘stealthily’, i.e., unobtrusively by the software that is the applied game, or by e.g. having the game’s facilitator keep the paraphernalia involved in gameplay. Several applied gaming case studies offer insights into the nature and usefulness of such behavioral data gathering and analysis [22, 23].

4 Discussion and Conclusion

This article has proposed a framework integrating validation in applied game design with an aim to extend the existing work on applied game validation. It further aims to contribute to the field of applied game design research and aid game design education. For many applied game designers, validation has been considered an afterthought. Playtesting and fine-tuning one’s design are seen as essential to the process. Different types of validity were introduced to broaden the scope of validation, making it encompass more than just standard evaluation trials. Our framework for applied game design validation was introduced as based on iterative design and design fidelity dimensions. The proposed framework links four types of design fidelity (specifications, concepts, prototypes, and integrations) to five types of validity (content, face, construct, concurrent and predictive validity), leading to eight tools and techniques that designers can choose to use in order to perform a validation study throughout their applied game design process. An overview of applicable validation tools and techniques was included with an aim to make validation more accessible to designers.

Limitations of this approach are as follows. Some of the proposed validation tools and techniques arguably come very close to verification (‘are we designing the game right?’) rather than validation (‘are we designing the right game?’). Those that come

too close to verification, e.g. a prototype stress test, were expressly omitted from the framework. It is also important to be aware that not all applied game design frameworks, heuristics or principles have been extensively validated, even though we refer to them in our validation framework. Further research should focus on this.

Additionally, this is not a quantitative validation benchmark which can be used in simulation or large-scale AAA companies. It is also not a large-scale validation process that validates the totality of any game. Although our framework is thus meant to be applicable to different types of game design, it favors an iterative, user-centered, co-design approach as well as the domains of education and health because of our experience with them.

Future research should test and validate the proposed framework. This is a first step towards creating a tool both accessible and useful for designers. Then again, some of the validation tools and techniques we have proposed are hardly ‘new’, but tried-and-true or even self-evident. The novelty of the framework lies in how it connects the design process to validity types, not necessarily in all the individual validation tools and techniques themselves.

Additionally, future research should focus on further testing the model amongst designers, as the question remains, are designers going to use this framework and is it feasible to ask this of designers whose main aim might not be to validate their applied game? As of yet, we do not know the answer. We are aware that the different paradigms behind our framework (particularly game design and social science) can be conflicting and confusing. The challenge is to integrate them in a feasible manner which does not hinder the designer in attaining both validity and a game. This article is an attempt at that integration, and a first step towards creating tools that lower the barrier to applied game validity.

References

1. Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., Berta, R.: Assessment in and of serious games: an overview. *Adv. Hum. Comput. Int.* (2013). Article no. 136864
2. Graafland, M., Dankbaar, M., Mert, A., Lagro, J., De Wit-Zuurendonk, L., Schuit, S., Schaafstal, A., Schijven, M.: How to systematically assess serious games applied to health care. *JMIR Serious Games* **2**, e11 (2014)
3. Mayer, I., Bekebrede, G., Hartevelde, C., Warmelink, H., Zhou, Q., van Ruijven, T., Lo, J., Kortmann, R., Wenzler, I.: The research and evaluation of serious games: toward a comprehensive methodology. *Br. J. Educ. Technol.* **45**, 502–527 (2014)
4. Sanchez, A., Cannon-Bowers, J., Bowers, C.: Establishing a science of game based learning. In: Cannon-Bowers, J., Bowers, C. (eds.) *Serious Game Design and Development: Technologies for Training and Learning*, pp. 290–304. IGI Global (2010)
5. Connolly, T., Stansfield, M., Hainey, T.: Towards the development of a games-based learning evaluation framework. In: Connolly, T., Stansfield, M., Boyle, L. (eds.) *Games-Based Learning Advancements for Multi-sensory Human Computer Interfaces: Techniques and Effective Practices*, pp. 251–273. IGI Global, Hershey (2009)
6. Kriz, W.C., Hense, J.U.: Theory-oriented evaluation for the design of and research in gaming and simulation. *Simul. Gaming* **37**, 268–283 (2006)

7. Feinstein, A.H., Cannon, H.M.: Constructs of simulation evaluation. *Simul. Gaming* **33**, 425–440 (2002)
8. Mueller, R., Thoring, K.: Design thinking vs. lean startup: a comparison of two user-driven innovation strategies. Presented at the 2012 International Design Management Research Conference, Boston, MA, 8 August (2012)
9. Kelley, T., Littman, J.: *The Art of Innovation: Lessons in Creativity from IDEO. America's Leading Design Firm*. Random House, New York (2001)
10. Kelley, T., Littman, J.: *The Ten Faces of Innovation: IDEO's Strategies for Defeating the Devil's Advocate and Driving Creativity Throughout Your Organization*. Random House, New York (2005)
11. American Educational Research Association, American Psychological Association, National Council on Measurement in Education: *Standards for Educational and Psychological Testing*. American Educational Research Association, Washington, D.C. (1999)
12. Saffer, D.: *Designing for Interaction: Creating Innovative Applications and Devices*. New Riders, Berkeley (2010)
13. De Freitas, S., Oliver, M.: How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Comput. Educ.* **46**, 249–264 (2006)
14. Marne, B., Wisdom, J., Huynh-Kim-Bang, B., Labat, J.-M.: The six facets of serious game design: a methodology enhanced by our design pattern library. In: Ravenscroft, A., Lindstaedt, S., Kloos, C.D., Hernández-Leo, D. (eds.) *EC-TEL 2012*. LNCS, vol. 7563, pp. 208–221. Springer, Heidelberg (2012)
15. Harteveld, C., Guimaraes, R., Mayer, I.S., Bidarra, R.: Balancing play, meaning and reality: the design philosophy of LEVEE PATROLLER. *Simul. Gaming* **41**, 316–340 (2010)
16. Harteveld, C.: *Triadic Game Design: Balancing Reality, Meaning and Play*. Springer-Verlag, London (2011)
17. Rooney, P.: A theoretical framework for serious game design: exploring pedagogy, play and fidelity and their implications for the design process. *Int. J. Game-Based Learn.* **2**, 41–60 (2012)
18. Dondlinger, M.J.: Educational video game design: a review of the literature. *J. Appl. Educ. Technol.* **4**, 21–31 (2007)
19. Charsky, D.: From edutainment to serious games: a change in the use of game characteristics. *Games Cult.* **5**, 177–198 (2010)
20. Annetta, L.A.: The “I’s” have it: a framework for serious educational game design. *Rev. Gen. Psychol.* **14**, 105–112 (2010)
21. Kato, P.M., Cole, S.W., Bradlyn, A.S., Pollock, B.H.: A video game improves behavioral outcomes in adolescents and young adults with cancer: a randomized trial. *Pediatrics* **122**, e305–e317 (2008)
22. Shute, V.J., Ventura, M., Bauer, M., Zapata-Rivera, D.: Melding the power of serious games and embedded assessment to monitor and foster learning. In: Ritterfeld, U., Cody, M., Vorderer, P. (eds.) *Serious Games: Mechanisms and Effects*, pp. 295–321. Routledge, New York (2009)
23. Shute, V.J., Ventura, M.: *Stealth Assessment: Measuring and Supporting Learning in Video Games*. The MIT Press, Cambridge (2013)