

How to Design Experimental Research Studies around Digital Badges

Rudy McDaniel
University of Central Florida
P.O. Box 161990, Orlando, FL 32816
rudy@ucf.edu

Joseph R. Fanfarelli
University of Central Florida
P.O. Box 161990, Orlando, FL 32816
joseph.fanfarelli@ucf.edu

ABSTRACT

Digital badging is becoming more popular in an assortment of disciplines, both academic and professional. Along with the success of practical badging initiatives, badging research is also moving at a rapid pace, a rate of change that may be intimidating to the uninitiated wishing to study them. However, there is a great need for additional research in light of the complexity of badging and the many contexts in which badging occurs. This paper outlines an approach to designing research studies around digital badges to assist researchers who are new to the field and looking to contribute. It begins by discussing how to form relevant research questions and how to approach the literature review, providing useful references as starting points. It then continues on to experimental design recommendations, discusses useful practices during experimentation, and concludes with recommendations for data analysis. Additionally, this paper describes challenges that are specific to badging and places them in context of the research design process. Multiple examples are provided to clarify these concepts.

Categories and Subject Descriptors

A.0 General – *Conference proceedings.*

General Terms

Measurement, Performance, Design, Experimentation

Keywords

Digital Badging, Education, Learning, Motivation, Goal Setting, Credentials, Assessment, Experimental Design, Research Design.

1. INTRODUCTION

Digital badging initiatives have gained traction in multiple professional domains and for a number of purposes. As evidenced by their use as digital credentialing technologies [1], reward systems [2], and as tools in other scenarios, badging is an emerging practice containing several dimensions ripe for study. However, good badging studies require researchers to make a number of careful decisions that can seem overwhelming for those new to the subject.

Copyright © by the paper's authors. Copying permitted for private and academic purposes.

In: D. Hickey, J. Jovanovic, S. Lonn, J.E. Willis, III (eds.): Proceedings of the Open Badges in Education (OBIE 2015) Workshop, Poughkeepsie, New York, USA, 16-Mar-2015, published at <http://ceur-ws.org>.

Currently, there is little support for new badging researchers looking to move into this area of scholarship. This paper aims to fill that gap by providing recommendations for planning and executing a badging study.

Badging research often spans multiple disciplines including psychology, computer science, educational technology, and the visual arts. As a result, some researchers from disciplines not trained in research methods may wish to learn basic strategies for conducting quantitative research so that we can more fully understand badge design and function, in an empirical sense, across academic boundaries. While researchers working in disciplines such as psychology and educational technology may already be well versed in experimental research design, those working in other fields may not. To address this issue and to provide context for the unique aspects of badging research, this paper situates badge-specific recommendations within the context of general good research practice. Overall, the purpose of this paper is to outline an approach for designing and developing badge-based research protocols based on our prior experiences developing and administering several such studies [3], [4], [5]. This is accomplished by discussing the research process as it relates to studies specifically designed around digital badges. Topics discussed include developing appropriate research questions for badges, considering how to develop dependent variables in these contexts, and recognizing the unique characteristics of badge-based data analysis.

2. DEVELOPING THE RESEARCH QUESTIONS

Badges are used for a variety of purposes, from incentivizing actions and behaviors to tracking performance outside normal channels of assessment and evaluation. They are also used in a variety of settings and for many purposes, from informal science learning in museums or wilderness scouting programs to formal coursework in the public school system and higher education.

Thus, when designing research questions for a badging study, it is useful to first consider the context and mechanics of the badging system. Before honing in on particular research questions, for example, several more general areas should first be considered. The following

procedural prompts are useful for thinking about how a research study might be generally framed, and then later operationalized, through more specific questions:

What is the context of the badging system? In other words, how is the badge system being used? Is it deployed in a children's museum for the purposes of informal science learning? Is the system integrated into a learning management system for the purpose of motivating undergraduate college students? Are badges being used in conjunction with a mobile fitness app to encourage healthy lifestyle choices? The overall context for the badging system is important to consider up front as a necessary precursor to the formation of specific research questions.

Who will design the badges? Will the experimenter design them? As an expert in the literature, the experimenter may be able to develop the most scientifically sound system. However, the experimenter is likely not as well-suited for design tasks such as graphic design or textual layout. In addition, design heuristics for badging research are still lacking. While past studies have developed some promising general guidelines, comprehensive blueprints for successful badge design for specific circumstances do not yet exist.

Is it necessary for the designer to have a relationship with the end users? While a researcher may understand the current state of badging knowledge, an educator may have a better understanding of the curriculum, the tasks with which her students are having difficulty, and knowledge about what motivates those students.

Who is awarding the badges? In other words, what is the central badging authority within a system? Badges can be provided for many reasons. They can be objective (e.g., a learner received a score of 90% or better) or subjective (e.g., a learner demonstrated exceptional effort on an assignment). The question of badge authority is especially important for subjective badges. Subjectivity, by definition, is ambiguous and open to interpretation. Thus, the perception of exceptional effort may differ by instructor.

There are a number of other guiding questions that are also useful for research design, but these are some of the more common prompts likely to be helpful across a variety of study types. The ways in which these questions are answered will influence the way research hypotheses are formed. As always, it is important to be clear and detailed in the formation of these experimental research questions. It is also important to be precise during the pre-experimental phase of the study in regards to the unique aspects of the badging system under investigation.

By formulating clear research questions at the onset, one can better determine the particular approach and instruments with which to design the study protocol. The research questions should take into account both the objectives of the research and the environmental factors in

which the research is occurring. They should additionally consider the audiences the investigators will have access to.

For example, in addition to the unique decisions made regarding the badging process, a study might be designed around the following research questions:

- How do badges affect motivation toward group projects in an online course focused on introductory psychology?
- Do students who earn more badges perform better in the course, as evidenced by earning higher grades?

The questions should be specific and measurable and may be explored through different types of research designs. In general, although the research questions may be refined after conducting the literature review, the overall purpose for the research should be decided upon at the beginning of the research process. This purpose will set the stage for the remainder of the protocol design.

In addition to the impact and influence of badges, the researcher might also identify the type of badging system they wish to study in the formulation of his or her research questions. For instance, if the aim is to compare open badge systems such as Mozilla's Open Badges to a proprietary badging system developed by a commercial vendor, the following research questions might be more appropriate:

- Are students more likely to consider badges useful when they can be permanently displayed outside of the course, even after completing their coursework?
- Are there differences in motivation or engagement toward course modules in students who use badge system A as compared to badge system B?

Another method of studying badges is in regards to their operational components. As Hamari and Eranti [6] explain, badges can be broken down into three primary components: a signifier, a completion logic, and a reward. These parts can each be considered separately within the design of a study to investigate badges using a finer degree of granularity. For example, here are two potential research questions focusing on separate components of badging according to Hamari and Eranti's framework:

- What visual signifiers are most effective for capturing the attention of players in the badges used within a racing-themed video game?
- How does the perceived degree of difficulty within the completion logic of puzzle-based games impact player enjoyment of those types of games?

There is a theoretically infinite number of potential research questions to be explored by badging studies. This flexible, purpose-driven research question generation process presents opportunities for badging studies to connect with or augment other research in areas such as sociology, psychology, or digital media and industrial design. For example, while badges are frequently thought of as rewards or credentialing systems, they can also

perform a number of other roles, serving as goal-setting mechanisms, social status indicators, and group identifiers [7]. Design cues can be tweaked as necessary to serve each of these purposes more readily and empirical data can suggest particular areas of the user interface in which to concentrate.

3. Literature Review

After the research questions have been formulated, it is time to plan out the specifics of the study. This cannot be effectively executed without a good working knowledge of the literature. As with any study, a thorough literature review is necessary to identify the conclusions of previous research and to discover what is not yet understood. New badging articles are being published frequently, so it is important to stay current with the latest developments in the field. In particular, annotated bibliographies focused on gamification (e.g., [8]) and digital badges (e.g., [9]) are valuable, as are studies outlining prior implementations of badges and the lessons learned through those experiences (e.g., [10]). Articles focused on the required and optional components and qualities of badges [6] can provide a deeper understanding of badges while also providing ideas for elements that can be manipulated during experimentation.

The diverse nature of badging, in combination with the frequency of new publications, can often make finding relevantly focused research difficult. Fortunately, the desired functionality of the system (e.g., reward or credentialing) has already been decided upon when formulating research questions. This can serve as a guide for appropriately narrowing the literature review from a large but broad set of results to a more narrow and precise body of work. The literature review might begin with research related to the broader focus of the research questions, such as badge motivation, reward, or credentialing. One needs to branch out to other relevant fields after this literature has been exhausted. For example, if badges are being studied as rewards with a specific emphasis on their potential to improve motivation, one may consider browsing the psychology-based literature to learn more about theories of motivation. Or, if credentialing is the focus, research on professional certifications may yield relevant knowledge. As digital badging is a fairly new field of study, it is possible that the specific research on digital badging related to the chosen topic may be sparse or even nonexistent. Considering broader areas or related disciplines of study will enable the discovery of knowledge that is potentially transferrable and may yield insights that prove valuable during hypothesis formation and experimental design.

4. EXPERIMENTAL DESIGN

After assimilating relevant background knowledge, experimental design can begin.

4.1 Experimental Manipulations

Since the specific badging functionality has already been selected by this point in the research design, identifying the experimental manipulations should not be overly taxing. Just as with any other experimental study, a control and experimental group will be necessary.

There are some unique concerns which should be thought through, however. For instance, if individual badges are being studied, the experimenter should record, without the participants' knowledge, who would have earned the badges had the badging system been implemented in the control. This will enable a comparison between those who completed the requirements for earning the badge of interest and received it against those who completed the requirements and did not receive it. Otherwise, it will not be possible to know if differences in the dependent variables between the control and experimental groups were due to the presence or absence of the badge itself, or due to some other factor associated with completing the requirements for earning the badge.

4.2 Demographics

Individual differences appear to be present in badging [7], [5], [11], but the results are not yet conclusive. It is important for future studies to continue to collect demographic data in order to understand how badges are perceived of and received by learners of varying genders and ethnicities.

Additionally, badging does include some novel demographic concerns that have implications for demographic survey design. For example, in response to the proliferation of badging in video games [12], [13], [14], video game players may have much more experience with badging than other participants. In games, badges may also be known as *achievements* or *trophies*. Since the effects of prior interaction with badging are not yet known, it can be helpful to include a question or instrument to assess the extent of prior interactions with games to aid the analysis of any unexpected results at the end of experimentation. Having this information will allow researchers to control for prior gaming experience within the sample during the data analysis phase.

4.3 Dependent Variables

A variety of dependent variables are relevant to badging studies, but some have received more interest than others. When badges are studied as rewards, intrinsic motivation [15] is commonly measured as an indicator of the participants' desire to complete the task simply for the reward of having completed the task. This is in contrast to extrinsic motivation which refers to the participants' desire to complete the task in order to gain some external reward such as money or a trophy. Intrinsic motivation is typically measured through the interest/enjoyment subscale of the intrinsic motivation inventory [16].

Learner engagement is another dependent variable that serves as an indicator of the participants' willingness to take an active role in the experimental task. This construct can be measured using questionnaires like the Classroom Survey of Student Engagement [17], the Student Course Engagement Questionnaire [18], or, for game and simulation-based studies, an engagement questionnaire [19]. Engagement can also be measured in terms of participants' level or frequency of activity on the experimental task. This might be indicated by the number of answers submitted, the number of times logged into the system, or the number of minutes spent on task.

Badges are frequently implemented to improve performance, so performance measures can also serve as dependent variables. The specific metrics used will depend heavily on the experimental task. In education studies, participants' final grades are typically appropriate data points to collect. Accuracy, as a ratio of correct to incorrect answers or classifications, is another popular metric that extends beyond academic environments. The researcher should consider which metrics best indicate success or failure in the environment being studied.

In some instances, badges themselves can be used as dependent variables. For instance, the number of badges earned may be a useful metric in correlational studies. Perhaps, as learners earn more badges, their perceived self-efficacy increases. Or, participants who earn more badges may exhibit more goal-directed behavior. If these questions are of interest to a researcher, they can be formulated as guiding research questions at the beginning of a research study's design (see Section 2).

5. SELECTING A BADGING SYSTEM

Selecting a badging system for an experiment is one of the most important tasks in designing a badging study. The chosen system will influence major aspects of the protocol including the risks of the study and the capabilities in badging experimentation. There are two primary routes: 1) Develop a system or hire a developer to develop the system; or 2) Use a commercial system that has already been developed.

5.1 Developing a System

Developing a badging system provides the greatest range of capabilities. When a system is developed from scratch, it can be built to exact specifications with the exact badging features one needs for a particular group of users. This is important because there is evidence to suggest that badging systems must be well-designed in order to fulfill their intended function [2, 4].

It is certainly possible to design and implement one's own system, but this may require skills outside of one's capabilities. Designing a badging system on one's own requires a number of diverse skills including programming, database design, graphic design, and instructional design. A background in writing or technical communication is also useful to determine the most effective methods for

structuring verbal and visual information within the context of badges.

When in-house developing prowess is insufficient, an external developer is needed to develop the badging system. In this scenario, it is obviously better to hire experienced developers if at all possible. With these professionals, finishing the system on time is more likely than it is in developing one's own system. Moreover, the same flexibility can be included in the developed system since it is still being developed from scratch to meet the specifications of the customer. If, however, development costs are high and budgets are small, hiring an experienced developer may not be an option. Or, financing the development of a very limited system may be the only possibility.

When working with a badge system developer, communication becomes very important. If the system's specifications are not well understood by the developer, errors may be made which could delay development or result in a system which does not meet the researcher's standards. Furthermore, bugs and specification discrepancies may not be discovered until experimentation has already begun, introducing variations that introduce minor confounds or even completely invalidate the data. This was the case with one of the authors' prior studies. They found out after the study had concluded that the participants enrolled in the non-badging section of a course were all emailed the same badge award notifications that the participants in the badging section received. This was problematic because the badges did not actually exist for those users enrolled in the non-badging section. A glitch in the badging system therefore led to problems with the separation between badging and non-badging sections of the course, leading to unreliable data.

It should be noted that either of these methods for developing a customized system comes at a cost. These two models provide the utmost flexibility for researchers, but they both require overhead that often makes this method of design impractical. It may take too long to develop a system that meets the needs of the experimenter. Or, it may be too expensive. In these instances, the enhanced flexibility may not be justified by the necessary resources to enable that flexibility. Further, if strict deadlines exist, attempting to develop a system in time to meet the study's goals will introduce a moderate level of risk into the study.

5.2 Using a Commercial System

When working with a tight budget or a strict timeline, a commercial system may be the only viable option. Here, the researcher will choose one of several commercially available badging solutions. These systems typically offer limited customization and may or may not require a fee, though this fee is typically very small in comparison to the cost of developing a new system. Perhaps the largest benefit of commercial systems is that they are already

developed and the majority of their bugs have already been corrected. These systems also offer good transparency in terms of features. The limitations and affordances are immediately apparent.

6. DURING EXPERIMENTATION

Experimentation in badging studies is very similar to other human-in-the-loop studies, but requires some special considerations. It is important for the experimenter to be aware of the special considerations of badge research and be prepared to record any unplanned events or unexpected observations regarding the badge system. Researchers administering badge-specific studies may want to pay particular attention to these three dimensions of complexity during the study.

6.1 Technical and operational complexity

As with any digital system, badging systems have the capacity to malfunction. This may happen due to problems within the software or usability issues with the participants. While the best course of action is to thoroughly test the system to avoid such issues before experimentation begins, bugs may still end up in the experimental system, potentially compromising the data. Once experimentation begins, the system should be regularly monitored and evaluated to identify any errors that arise. For instance, a participant may mention receiving a badge that they did not earn, or not receiving a badge that they did earn. If the experimenter decides that the participant is correct, the malfunction is obvious and may be quickly remedied (after a note is made to evaluate the impact of this confound). However, participants may not be familiar with the criterion for earning a badge, or the criterion may be too subjective for the participant to make a confident decision. As a result, the experimenter should routinely check to ensure that the badges are being awarded in the way that was designed. This can ensure that errors are caught early and can be accounted for during data analysis.

6.2 Behavioral complexity

Participant behaviors in these types of studies can be very informative. Especially if the experiment only lasts for a short duration and the experimenter and participants are co-located, it is helpful to record behavioral observations. Noticing behaviors that are indicative of changes in dependent variables can provide useful information for data interpretation or ideas for future experimentation. For example, participants shouting in frustration or satisfaction may be showing indicators of engagement or immersion. Or, if the badging system uses notifications that interrupt the task whenever a badge is awarded, and a student groans and tries to quickly click out of the notification every time it appears, this could be an indicator that the notification system is invasive. A future study may wish to see if these notifications are detrimental to the badging system's effectiveness. Participants may mention how much they enjoyed or disliked a particular aspect of the badging experience (e.g., "The badges were hideous. You should

really hire a graphic designer"). This provides a better understanding of whether the results were due to the inclusion/exclusion of badges, or due to some unconsidered extraneous variable such as visual design.

6.3 Temporal complexity

The experimenter may also wish to examine badge access patterns. This creates challenges in regards to time. What is the appropriate duration in which to consider the earning of badges, and what sort of longitudinal impact will the acquisition of these badges have for participants? Further, how will badges be made available to earners on a long term basis? When badges can be viewed on demand, differences in the number of times the badges were accessed in a particular experimental session may be indicative of engagement or interest. A user who exerts additional effort to frequently view badges or a list of possible badges that can be earned is probably more interested in the system than a user who never or rarely accesses the badges.

In sum, researchers are still trying to understand how to build better badges. An understanding of what factors are important to their effectiveness and which areas of complexity are most relevant to digital badging studies will help to further this effort. Factors related to these areas of complexity may arise unexpectedly during experimentation and should be recorded.

7. DATA ANALYSIS

During data analysis, it is important to consider who actually earned and did not earn each badge. If badges could be earned multiple times, it is also important to consider the number of times each badge was earned. By doing this, the researcher can test for effects of the inclusion, earning, or award of badges on an individual level, identifying the impact of badges independently instead of just observing the system which contains them. These analyses could yield results that help badging system designers understand which criteria are the most useful for badges in their system.

Also, simply having a badge available in a course is unlikely to be sufficient as a manipulation if the badge is never earned or seen. For example, consider the case of hidden badges, or badges users do not know about until they are earned. If these hidden badges are never earned, they serve the same role as if they never existed; they are invisible to the user. The fact that the badges technically could have been earned are unlikely to have affected the dependent variables. In other words, the relationships between badges can be more complex than they seem at first glance. Badge visibility should be considered carefully during data analysis. At minimum, be sure to make the distinction between:

- Badges that were available, but never seen
- Badges that were available, and were seen, but not earned

- Badges that were available, seen, and earned once
- Badges that were available, seen, and earned multiple times

Breaking down the results in this manner will enable formation of more specific conclusions regarding the data.

8. CONCLUSION

While general rules and strategies that pertain to experimental design are often applicable to badging experiments, badging presents some novel challenges that require careful consideration. It is important to understand these challenges because more research is needed in this area. Badging research is still in its infancy, despite the rapid growth it has recently experienced. Although recent years have generated exciting insights and ideas about digital badging, future research will continue to illustrate the precise conditions in which digital badges thrive. Sound research design will help us to design the experiments that collect the empirical data that help us to outline these conditions. Meticulous experimentation will yield data that will inform practitioners and future researchers alike, enabling higher quality research and more effective badging implementations.

9. REFERENCES

- [1] Carey, K. 2012. Show me your badge. *The New York Times*. Retrieved from <http://blogs.maryville.edu/kfletcher/files/2012/11/Show-Me-Your-Badge-NYTimes.pdf>
- [2] Blair, L. 2012. The use of video game achievements to enhance player performance, self-efficacy, and motivation. *Doctoral Dissertation*, University of Central Florida.
- [3] McDaniel, R. and Fanfarelli, J.R. in press. A digital badging dataset focused on performance, engagement, and behavior-related variables from observations in web-based university courses. *British Journal of Educational Technology*.
- [4] Fanfarelli, J. 2014. The effects of narrative and achievements on learning in a 2D platformer video game. *Doctoral Dissertation*, University of Central Florida.
- [5] McDaniel, R., Lindgren, R., and Friskics, J. 2012. Using badges for shaping interactions in online learning environments. In *Proceedings of IPCC* (Orlando, FL, USA, October 8 – 10, 2012).
- [6] Hamari, J. and Eranti, V. 2011. Framework for designing and evaluating game achievements. In *Proceedings of DiGRA* (Hilversum, Netherlands, September 14 - 17, 2011).
- [7] Antin, J. and Churchill, E.F. 2011. Badges in social media: A social psychological perspective. In *Proceedings of CHI* (Vancouver, BC, Canada, 2011).
- [8] Hamari, J., Koivisto, J., and Sarsa, H. 2014. Does gamification work? A literature review of empirical studies on gamification. In *Proceedings of the 47th Hawaii International Conference on System Sciences* (Hawaii, USA, January 6 – 9, 2014).
- [9] Grant, S. and Shawgo, K.E. 2013. Digital badges: An annotated research bibliography. Retrieved from <http://hastac.org/digital-badges-bibliography>.
- [10] Grant, S. 2014. What counts as learning: Open badges for new opportunities. *Digital Media and Learning Research Hub*. Irvine, CA. Retrieved from http://dmlhub.net/sites/default/files/WhatcountsAsLearning_Grant.pdf
- [11] Montola, M., Nummenmaa, T., Lucero, A., Boberg, M., and Korhonen, H. 2009. Applying game achievement systems to enhance user experience in a photo sharing service. In *Proceedings of Mindtrek: Everyday Life in the Ubiquitous Era* (Tampere, Finland, September 30 – October 02, 2009).
- [12] Microsoft. 2015. Xbox One achievements and challenges. *Xbox One Support*. Retrieved from <http://support.xbox.com/en-US/xbox-one/games/achievements-and-challenges>
- [13] Sony. 2015. Get trophies. Get recognition. *Playstation Community*. Retrieved from <http://us.playstation.com/community/mytrophies>
- [14] Steam. 2015. Game stats and achievements in steam. *Steam Support*. Retrieved from http://support.steampowered.com/kb_article.php?ref=4462-UIOB-0425&l=English
- [15] Ryan, R.M. and Deci, E.L. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*. 55(1) (2000). 68-78.
- [16] SelfDeterminationTheory.org. n.d. Intrinsic motivation inventory (IMI). *Self-Determination Theory: An Approach to Human Motivation and Personality*. Retrieved from <http://www.selfdeterminationtheory.org/questionnaires/10-questionnaires/50>
- [17] Ouimet, J.A. and Smallwood, R.A. 2005. Assessment measures: CLASSE–The Class Level Survey of Student Engagement. *Assessment Update*, 17(6), 13-15.
- [18] Handelsman, M.M., Briggs, W.L., Sullivan, N., and Towler, A. 2005. A measure of college student course engagement. *J. Educational Research*, 93(3), 184-191.
- [19] Charlton, J. and Danforth, I. 2005. Distinguishing addiction and high engagement in the context of online game playing. *Computers in Human Behavior*, 23(3), 1531-154.