Experiments in design research: artificial and naturalistic settings for the assessment of user experience

Experimentos em pesquisa de design: configurações artificiais e naturalísticas para a avaliação da experiência do usuário

Leandro Miletto Tonetto
PhD in Psychology (Pontifícia Universidade Católica do Rio Grande do Sul) and Post-Doc in Healthcare Design (Georgia Institute of Technology-GeorgiaTech) Graduate Program in Design, Universidade do Vale do Rio dos Sinos email: ltonetto@unisinos.br

Valentina Marques da Rosa
Master’s Degree in Design (Universidade do Vale do Rio dos Sinos) Department of Industrial Engineering, Universidade Federal do Rio Grande do Sul email: valentina.rosa@hotmail.com
Filipe Campelo Xavier da Costa
PhD in Management (Universidade de São Paulo)
Graduate Program in Design, Universidade do Vale do Rio dos Sinos
email: fcampelo@unisinos.br

Fabrício Farias Tarouco
PhD in Communication Sciences (Universidade do Vale do Rio dos Sinos)
Graduate Program in Design, Universidade do Vale do Rio dos Sinos
email: ftarouco@unisinos.com
ABSTRACT

This paper discusses the development of experiments to assess user experience with innovative artefacts, based on academic and market research. It discusses ecological validity in the use of naturalistic experiments and when artificial experiments are “good enough”. Due to its potential to keep the study’s internal validity high, the artificial scenario is the first choice. When artificial settings threaten the study’s external validity, naturalistic experiments are better choices. Design research requires flexibility when planning experiments, since the usual choices made in other sciences might not be the best ones to be made.

KEYWORDS

RESUMO

O presente artigo discute o desenvolvimento de experimentos para avaliar a experiência do usuário com artefatos inovadores, tendo como base pesquisas acadêmicas e de mercado. Discute validade ecológica no uso de experimentos em ambientes naturais e quando experimentos artificiais são “bons o suficiente”. Devido a seu potencial para manter a validade interna dos estudos alta, o ambiente artificial é a primeira escolha. Quando ambientes artificiais ameaçam a validade externa, experimentos em ambientes naturais são a melhor escolha. A pesquisa em Design requer flexibilidade no planejamento de experimentos, pois as escolhas usuais de outras ciências podem não ser as melhores escolhas.

PALAVRAS-CHAVE
1 Introduction

According to Steffen (2013), the term experiment refers to systematic methods used by science, intending to increase the knowledge regarding something specific, based on statistical principles. Günther (2006) understands experiments as one of the main ways of practicing science in empirical social sciences. However, the term gradually stopped being used only in relation to sciences, covering other areas of study, such as literature, theater, films, music, arts and design (Steffen, 2013).

In the experimental sciences, it is not always common to develop experiments in natural settings (Falk & Heckman, 2009). They are often performed in laboratories, as they make it easier to control unwanted interferences that could threaten the validity of the results (internal validity), such as other consumers advising research participants on which product to buy, time and temperature, etc. The artificial setting, often seen in experimental settings, is frequently criticized, based on the idea that the research results would not reflect the reality outside of the laboratories (ecological validity) (Christensen, 1989).

Experimental studies in artificial settings are commonly understood as more desirable than research in natural settings. However, in applied areas, such as Design, studies may require different approaches than other areas (e.g. experimental psychology, concerning the setting of its experiments).

User experience research for the development of innovative products and services often requires the measurement of variables, such as the perception of users in research that tests a variety of new versions of products and services (Tonetto, Brust-Renck & Stein, 2014). Mettler, Eurlrich and Winter (2014) describe that experiments help verify if an artifact is superior to something specific, identifying ways to improve products and design processes. Seeking to understand more about how to work with the complexity related to experimental research and creative design methods, Mainsah and Morrison (2013) created a manifesto, in which they state that the area still needs to learn ways to deal with problems in real settings.

Experiments in design research are often misinterpreted. Researchers like Gaver (2012) believe that their focus originated in the exact sciences, such as engineering. For Mettler et al. (2014), on the other hand, the existing experimental protocols in Design are based on aspects arising from behavioral research, which is not always specific or sufficient enough for research in the area.

More than the simple task of working on the reproduction of predefined methods, in this article, it is considered as part of the designer’s role to
understand how it is possible to use research in artificial and natural settings in Design, specifically concerning user experience. In this article, we seek to understand how to plan design experiments on user experience with innovative artifacts. To this end, it explores the understanding of two main questions: “The ‘problem’ of ecological validity and the use of experiments in natural settings” and “when are experiments in artificial settings good enough?” The study carried out can be considered a scope review, not adopting quantitative criteria, but of relevance, for the selection of the papers included in the analysis. Before presenting the analysis performed (chapter 3), the fundamentals of experimental research are discussed in the following section.

2 Experimental research in Design

In Design, independent variables commonly manipulated in experiments are related to the characteristics of products, such as packaging, color and texture. Regarding the user’s experience, the dependent variables are those that are measured to verify the impact of the experimental manipulation, such as emotion, behavior or perception (Christensen, 1989). An example could be the manipulation of the color of a package to study if cool colors, instead of warm colors, can impact consumer’s buying behavior in different ways.

Independent variables can be manipulated between or within groups (Schweigert, 1994). Changing the color of a product can be an example of manipulation between groups, since each level of the independent variable (color, in this case) can be presented to different groups of people to neutrally evaluate the effect of the color change. This type of manipulation avoids conscious comparisons between alternatives.

Another example is to test whether the use of an innovative material in a shoe would produce better levels of perceived comfort. If the researcher asks people to compare products, they may tend to evaluate diverse materials in distinctive ways, not because they would be assessed differently in a natural situation, but because the researcher asked participants to compare them. A biased result could lead the company to invest large amounts of resources to change materials without a real reason. If this experiment focused on shoe color preference, it would be better if the independent variable were manipulated within groups, which means that the same group of people, instead of evaluating a product, would compare more than one (or choose between them).
When the researcher has defines which variables will be manipulated (independent variables) to measure their effects on other variables that will be measured (dependent variables), it is common to perceive that some of them can threaten the quality of the results, if they are not measured or controlled, such as gender, age, or previous habits (Schweigert, 1994). For example, in the experiment in which the color of a product was manipulated, the preference for warm or cool colors can be related to the user’s previous habits, such as the tendency to buy a specific brand of the product, which uses a lot of red color. This habit can lead the consumer to an automatic response, which would be to choose the warm colored product due to an automatic behavior. To plan the mentioned experiment, the researcher needs to know the market, identifying which variables and habits are involved in the evaluated phenomena. In this case, an option could be to test the product among people who do not have the habit of using or buying the same brand.

Experimental research is related to the investigation of causality relationships (Collins, Joseph & Bielaczyc, 2004; Pinheiro & Günther, 2008). In the aforementioned example, the person assumes that a color can be responsible for preference. Imagine that the product is blue and that you want to test, based on the results of sales of other products, whether an orange or red version can cause better sales results. Choosing a supermarket as a “laboratory”, it would be possible to define three days a week when the product shows similar sales results and test it under the three conditions (blue, orange and red), recording the volume of sales after the manipulation to compare the differences between before and after changing the color. It is common to measure the dependent variable (in this case, sales) before and after manipulation (Christensen, 1989) to assess its variation.

Another example related to shopping experience would be to investigate changes in the user’s emotional experience (dependent variable) when shopping in a silent retail store as opposed to a loud one (independent variable), illustrated in Figure 1. In this case, it would be necessary to record users’ emotions (through a survey, for example) before and after the shopping experience.
Cash (2011) suggests that the most common problems in design experiments are (a) evaluating the context: usually related to the lack of adequate definition, leading to difficulties in understanding, for example, the issues behind the resolution of a task, such as a consumer choosing a product; (b) understanding the system: some underlying variables under test are often unknown or not controlled in the studies, which leads to a lack of reliability in the results; (c) implementing the method: standardization and control are often a methodological issue; and (d) controlling and normalizing: there is a lack of placebos and control groups to allow baseline measures.

How can we design experiments to identify causal relationships, avoiding the problems mentioned by Cash (2011), and maintaining reliable results, when trying to understand the reality of controlled experimental situations? This is the question that will be answered in the following section.

3 Analysis and Discussion of Results

This chapter discusses the importance of ecological validity in studies in natural settings and the benefits of experiments in artificial settings. Thus, we sought to understand the context in which both can be applied.
3.1 Ecological validity in experiments in natural settings

The results of experiments can be questionable to understand real situations, if the setting in which the experiment is carried out is so artificial that it does not reflect life outside of the laboratory. When this problem occurs, the experiment has a problem related to ecological validity (Schweigert, 1994).

The term “ecological validity” often refers to the relationship between real-world phenomena and the investigation of these phenomena in experimental contexts (Schmuckler, 2001). Researchers are usually concerned with internal validity, but not with ecological validity, since most prototypes of innovative projects are not fully functional. In other words, they are usually concerned with how well prepared the experiment is, offering reasonable certainty that the effect measured was caused by experimental manipulation (Lew, Nguyen, Messing and Westwood, 2011). Thus, by increasing realism, the construct's validity, which refers to the correlation between a measurement scale and the theoretical/scientific construct, is also increased, since the users’ responses are closer to what they would be in real-world situations.

There are at least three dimensions involved in ecological validity to be considered by the designer in experimental research (Schmuckler, 2001):

A. The setting in which the research is conducted: the use of artificial settings has been discussed among the scientific community, suggesting that researchers should pay more attention to the nature of the investigation itself, focusing their work on the situations/circumstances in which the phenomena in study occurs. There is a reciprocal relationship between the person and their social context (Günther, 2009; Seizel's (2016). Some changes may be necessary in the way researchers think about experiments to maintain the integrity of social and cultural real-life situations (Schmuckler, 2001).

B. The nature of the stimuli used in experiments: the use of intangible, discontinued and only slightly real stimuli, such as those often observed in the early stages of innovative projects, differ critically from those found in real-world situations (Schmuckler, 2001). By not representing real life, applying research results may be dangerous, due to low ecological validity.

C. The nature of responses from research participants: the potential problem related to ecological validity is also associated with artificiality. If the response from users is not natural, representative or appropriate to represent
real behavior, the researcher may face a problem of ecological validity.

For example, if researchers want to know if manipulating the pattern of a plate can change people’s perception of a meal, they can think of using images in their experiments as stimuli, rather than real plates. This type of practice is common in the area (Okajima & Spence, 2011). It would be easy, when using digital images, to switch plates, keeping the appearance of the meal. This means that researchers would be successful in manipulating only the plate, without having to change anything else in the image. The internal validity of the study may be high (which means that the experimental procedure itself has no problems and that the independent variables are adequately manipulated), but are these results useful in the real world?

The obvious problem in this example is that people do not eat digital images. Their perceptions may not change, since the experimental situation is very artificial and distant from the real experience. Users’ responses may have been different if they were hungry, waiting at a restaurant and paying for the meal.

The difficulty in manipulating some real-life conditions can even reduce the interest in the development of experimental research. This is not always the case, as some experiments can be easily adapted to better reflect real situations, increasing their ecological validity. For this purpose, an experiment in a natural setting was developed in an office by one of the authors of this article (unpublished; a non-Disclosure Agreement was signed). A company carried out a research to evaluate the effect of a new type of foam on the comfort of an office chair. The researcher used four real offices where people sit for most of the day, performing activities in front of computers. The conditions of the experiment can be found below.

A. In the first two offices, employees were provided with the usual chair (control group), which was used for one week. At the end of this week, they answered a survey about their perception of comfort. The following week, they were provided with new chairs, and at the end of five working days, they responded to the same survey.

B. In the two other offices, employees received the new chair first, and the normal one afterwards (the opposite if compared to the first two offices). This procedure was carried out by the researcher to control the effect of the suggestion that the two chairs under test could be different,
even if they had the same appearance. Figure 2 illustrates the conditions of the experiment.

Figure 2: Experiment that tests the comfort of different types of foam in the same chair.

In each of the offices, between 30 and 40 people participated in the research. Comfort was measured using five-point self-assessment scales. The results were compared and did not present differences in the employees’ perception of comfort. This result illustrates the idea of Mettler et al. (2014) regarding the relevance of design experiments. For them, this is one of the best methods to observe the usefulness of a new artifact, putting it to use. In this case, the product was not launched, avoiding large sums of money to be spent.

Naturalness and artificiality, discussed in depth in the next section, are not necessarily understood as a dichotomy. The experiment described below (unpublished; market research), researchers focused on the effect of manipulating LED lighting in clothing stores (independent variable) on people’s self-assessment of desire (dependent variable). Firstly, in-depth interviews were conducted to understand how people assess lighted environments that cause desire, in their perception. Digital images, in this case, were used to help provide stimuli to the individuals during the interviews. Based on the results, a set of environments was designed in real retail spaces, where people had a shopping experience, without actually buying any product. All participants had the exact same experience with the space. At the end of the visit to the real store, they were asked to complete an experience survey. The analysis provided
a comparison between the different environments, helping designers understand how to evoke desire through the use of light.

This experiment suggests that, instead of questioning the real value of experimental methods in design research, the concern with the artificiality of experimental settings should inspire researchers. The important question to be answered is how to better represent real design problems in empirical contexts (Schmuckler, 2001), as discussed below.

3.2 Benefits of experiments in artificial settings

The difference between experimentation and observation lies in the researchers’ ability to manipulate the cause of a phenomenon and control the test setting. In controlled settings, only independent variables are subject to deliberate manipulation (Mettler et al., 2014). The reasons for developing artificial experiments are (a) sometimes it can be difficult to control the interference of external variables in real-world situations (Christensen, 1989), although product testing is crucial to assess the design, and (b) it is common to have problems that are easy to solve, which means that they are relatively independent of external variables. These reasons are explored as follows.

Some naturalistic experiments would have to be planned in very complex settings to be representative of real life (item “a”). As an example, Tonetto, Klanovicz and Spence (2014) sought to investigate the effect that the manipulation of sounds produced by high-heeled shoes have on women. This study was related to how they reported their levels of valence (pleasure), arousal and dominance, as well as any changes in a variety of body sensation measures. All perceptions were collected using self-assessment scales. The study had a sample of 48 women in an artificial setting, since the context of real settings, such as shoe stores or everyday situations, would make it difficult to perform the experiment, due to the presence of music, other people and other external factors that could interfere with the results. Only one group of women participated in the study. Unlike a significant part of the experiments, manipulation of the independent variable in a single group was desired, since, in purchase situations, consumers try on and compare different shoes in a store, thus, the experimental situation simulates reality. The women walked on a “virtual runway”, in a laboratory, while listening to four interaction sounds each time their feet touched the floor (high heels with leather or polypropylene soles in contact with ceramic floors or carpets), as shown in Figure 3.
Before starting to walk, each participant answered a survey with questions regarding weight, height and age, and personal style. Afterwards, they were asked to complete another survey, in which they explained how they felt listening to the sounds, describing the sensations they experienced. After analyzing the data, it was possible to notice that the type of sound interaction impacted the users’ levels of valence (pleasure), arousal and dominance, as well as their bodily sensations.

Leaving the universe of physical artifacts, the experiment used here as an example focuses on the experience of users with beauty products. The research aimed to understand the effect the size of the choice set (colors and shapes of packaging - independent variables) on the experience of buying products (dependent variable) (Assis and Tonetto, 2016), since the literature indicates that large sets of options may lead to difficult in making choices (Iyengar and Agrawal, 2010).

Data collection was conducted in a virtual setting, where users would find a set of choices to be made, instead of using a real retail space. The artificial setting promised to be more successful, as it avoids distractions with other products, consumers and brands. Even though it is artificial, it is somehow naturalistic in terms of the size of the choice set. The average size of set was investigated and determined. From this number, two experimental groups were planned: one with 50% of the average choice set for each of the evaluated products, and the other with 150%, leading to three groups (a control group with the average size of the choice set, and the two experimental groups presented). The total
sample (96 people) was divided in the three groups, and the experience was evaluated using seven-point scales. The analysis revealed that the largest set of options provided a more positive experience.

As it would be in this case, the high cost of the research is commonly an obstacle to conduct experiments in natural settings. The cost is often associated with prototyping innovative artifacts, or with the development of commercial settings to be used as laboratories.

Furthermore, the increase in realism may cause a reduction in the experimental validity by losing some of the control over the experimental situation (Lew et al., 2011). In the example about the sound of shoes, a real store setting would expose people to a variety of stimuli that could affect their assessments, such as noise, interaction from other customers, etc. Although an experiment in a natural setting may have a higher level of ecological validity, its internal validity would be weak, making it more interesting to perform the study artificially.

Simple problems can be tested in artificial settings with little or no impact on the quality of results (item “b”), when they are not directly related to external interference. An example can be seen in one of the experiments carried out by Rosa, Spence and Tonetto (2018), which tested the impact that changes in colors and shapes of food packaging have on the user’s experience. Based on other studies, the authors assumed that past experiences tend to influence the perception of new artifacts. To avoid external interference, such as memory, images of packages without specific products and brands were presented to participants. In this case, the aim was to isolate perceptions solely about the food packaging.

Thus, sometimes artificiality is desirable. It is important to highlight that it is risky to increase realism in experimental conditions. It introduces disturbing stimuli into the experimental situation, such as the music referred to in fashion retail, reducing the control of the experimentalist over the measured variables (Lew et al., 2011). Additionally, Kvavilashvili and Ellis (2004), reviewing nearly twenty years of discussions regarding the controversial battle between traditional and ecological approaches, suggest that the main concerns in the area are generalization and representativeness (the extent to which a study can be developed corresponding to the occurrence of a phenomenon in everyday life). Therefore, artificiality is related to representativeness, while the potential applications of research results are related to generalization. This is the reason why artificial experiments are sometimes good enough to be adopted in design research.
4. Conclusions

Traditionally recognized as an interdisciplinary area, Design presents many possibilities and challenges in the development of experiments. By comprehending real cases, the reader can understand why experimental research has the potential to play an important role in studies on user experience, although it is sometimes necessary to adapt “purely” experimental approaches to better answer research problems in the area.

It is not correct to think that experimentalists plan studies that necessarily produce artificial results or are not applicable in contexts outside of the laboratory. Experiments can be planned in natural settings, but they require the attention of experimentalists when designing them. The researcher must comprehensively investigate the variables that may be worth measuring, in addition to verifying that the experimental setting does not threaten the validity of the results. The use of a naturalistic approach is not justified if the internal validity of the experiment is low.

On the other hand, sometimes artificial experiments are good enough, since the internal validity is generally higher than in other scenarios and the independent (or experimental) variables are easier to control. Not only can artificial experiments be adequate to answer the research aims, but they are commonly associated with lower implementation costs. In research on user experience with innovative artifacts, fast and low-cost experiments minimize the potential negative impacts of “failures” to make prototyping and testing processes recurrent, as adopted by several companies.

We believe that this paper provides enough information to state that artificial settings, in several cases, may be the first choice when planning innovative experiments on user experience, due to their greater potential of maintaining the internal validity of the study high. If the artificial setting threatens the external validity of the study, then a naturalistic experiment based on the real world may be more interesting. It may not be worth having high external validity if the price paid is low internal validity.

References
ASSIS, M. V.; TONETTO, L. M. Questioning the choice overload effect through design research. In: Proceedings of the Tenth International Conference on Design and Emotion, 2016, Amsterdam. Anais...


---

**Leandro Miletto Tonetto**

PhD in Psychology (Pontifícia Universidade Católica do Rio Grande do Sul), Post-Doc in Emotional Design (Delft Institute of Technology) and Post-Doc in Healthcare Design (Georgia Institute of Technology - GeorgiaTech). Professor of Design, School of Creative Industries, Universidade do Vale do Rio dos Sinos. Leandro Miletto Tonetto studies design for well-being as a means to foster human development and health. He is co-founder of WB Wellbeing Consultancy and Zooma Consumer Experience. He is a board member of the Special Interest Group in Wellbeing, Happiness, and Health (SIGWell) of the Design Research Society. His research is funded by CNPq [Grant # 307113/2017-5].

TONETTO, L. M.; PEREIRA, A. S.; KOLLER, S. H.; BRESSANE, K.; PIEROZAN,


Valentina Marques da Rosa

Master’s degree in design (Universidade do Vale do Rio dos Sinos). PhD Candidate, Department of Industrial Engineering, Universidade Federal do Rio Grande do Sul. Valentina Marques da Rosa is PhD student in Industrial Engineering at Universidade Federal do Rio Grande do Sul, Brazil, and co-founder of WB Wellbeing Consultancy. She studies the relationships between users and artefacts to enhance users’ wellbeing and health.


Filipe Campelo Xavier da Costa

PhD in Management (Universidade de São Paulo), Master in Science in Management (Universidade Federal do Rio Grande do Sul) and Master in Strategic Design (Universidade do Vale do Rio dos Sinos / PoliDesign Milano). Professor of Design, School of Creative Industries, Universidade do Vale do Rio dos Sinos. Filipe Campelo Xavier da Costa studies the relationships between user experience and strategic design processes,
design for behavioral change, and design processes and practices in food design. He is a board member of the Red Latinoamericana de Food Design (LAFD).


**Fabrício Farias Tarouco**

PhD in Communication Sciences, Master in Strategic Design (Universidade do Vale do Rio dos Sinos) and Advanced Studies in Lenguajes y Sistema Informátivos (Universidad de Zaragoza). Professor of Design and Architecture, and Coordinator of the Bachelor of Design, School of Creative Industries, Universidade do Vale do Rio dos Sinos. Fabricio Farias Tarouco has experience in Territorial Design, Communication Metropolis, Creative Cities, Smart Cities, Tourist Strategies, Urbanism and Experience in the Public Space. He worked as an AD HOC Consultant at the Secretariat of Science, Innovation and Technological Development of Rio Grande do Sul for Creative Industry projects.


Received: April, 2020
Accepted: June, 2020