How Wild Is Too Wild: Lessons Learned and Recommendations for Ecological Validity in Physiological Computing Research

Elise Labonte-LeMoyne, François Courtemanche, Marc Fredette and Pierre-Majorique Léger
Tech3Lab, HEC Montreal, Montreal, Quebec, Canada

Keywords: Ecological Validity, In-the-Wild, Representative Design, Physiological Computing.

Abstract: While many call for increased ecological validity in physiological computing research, implementing very naturalistic studies can be challenging. In this paper, we present a way to quantify ecological validity to allow comparisons between studies. We also present a critical look at four types of studies that have emerged from quantifying the ecological validity of our past experiments. Finally, we provide recommendations and lessons learned from our own work conducting studies that span a wide range of levels of ecological validity for researchers who wish to do more in the wild research.

1 INTRODUCTION

It has been argued for some time now that a person's emotion are intimately tied to the context in which they are experienced (van den Broek, 2012).As our technology use has become ubiquitous, it is difficult to justify how a laboratory study where the subject is properly seated with their head strapped to a chin rest can generalize its conclusions to a day to day context. Finding balance between ecological validity and reliability is a constant challenge. Fortunately, technology has improved greatly in recent years and usable data can be acquired more and more in real world settings.

In 2009, van den Broek highlighted the need for clearer guidelines on how to conduct more ecologically valid physiological computing research (van den Broek, Janssen and Westerink, 2009). A decade later, we present a method to quantify ecological validity that can be used to compare studies to each other. We also present some lessons learned and recommendations from our own experiences with a variety of levels of ecologically valid experiments to help other researchers in this community be successful in their endeavours.

2 ECOLOGICAL VALIDITY AND PHYSIOLOGICAL COMPUTING

Ecological validity has been defined in many ways as it often overlaps with other concepts such as generalizability, representative design, realism and external validity (Kieffer, 2017). While initially the term referred to a vision research concept (Brunswick, 1949), since the mid-1970s it has been commonly used according to the definition by Bronfenbrenner: “Ecological validity refers to the extent to which the environment experienced by the subject in a scientific investigation has the properties it is supposed or assumed to have by the investigator.” (Brofenbrenner, 1977). In essence, it refers to how the research context is representative of the real-life situation the results should be generalized to.

Applied research, such as physiological computing, strives to have the highest possible ecological validity and to get away from the intimidating white coat researchers in traditional psychology. As Bronfenbrenner put it, laboratory studies can sometimes be “the science of (...) strange behaviour in strange situations (...) for the briefest possible periods of time” (Brofenbrenner, 1977). As such, conclusions from this type of research can sometimes provide less insights that expected to inform the advancement of technology and human-computer interaction (HCI). This is associated with a duality mind-set in some communities that often contrasts laboratory experiments as “bad” and not particularly valid, while field research is considered “good” and much more valid. However, we argue that the concept of ecological validity should be considered as a continuum and the degree needed for a particular study should be determined with great

Labonte-LeMoyne, E., Courtemanche, F., Fredette, M. and Léger, P.
How Wild Is Too Wild: Lessons Learned and Recommendations for Ecological Validity in Physiological Computing Research.
DOI: 10.5220/0006962901230130
In *Proceedings of the 5th International Conference on Physiological Computing Systems (PhyCS 2018)*, pages 123-130
ISBN: 978-989-758-329-2

care as it can often be accompanied by a trade-off in other aspects of the methodology and in data quality.

3 QUANTIFYING ECOLOGICAL VALIDITY

While there are often calls for increased ecological validity, rarely do we see guidelines on how to do so. To improve ecological validity, we first need a way to quantify it in order to compare studies to one another. To our knowledge, only two attempts to operationalize ecological validity have been published, one in developmental psychology (Schmuckler, 2001) and one in Human-Computer Interaction (Kieffer, et al., 2015) that built upon the work of Schmuckler and added dimensions regarding the technology being used. Thus we will move forward with the ECOVAL framework proposed by Kieffer and adapt it slightly for use in physiological computing.

The ECOVAL framework (Kieffer, et al., 2015) is composed of 6 dimensions. Environmental signals and objects that refer to the environmental context, Test medium and User interface that refer to the system employed, Task and Behaviour that refer to the participants themselves. Each dimension can be rated as low, medium or high ecologically valid (see Table 1 for the definitions of the dimensions and their levels). For the purposes of physiological computing, we propose the addition of a 7th dimension that refers to the reactivity to the measurements employed. As this table was initially developed for user testing, some of the parameters may seem less applicable to physiological computing (using paper mockups). We chose to leave the original dimensions and levels, as they were validated by the original authors. Future research could adjust these for physiological computing and validate them properly.

The term reactivity to measurement comes from Goodwin (Goodwin, et al., 2008), who referred to the intrusiveness, or the impact of the measurement processes on the research subject's behaviour, as a separate concept to ecological validity. He considered ecological validity, repeated assessment and reactivity to measurement as "key issues relevant to behavioural assessment strategies in the behaviour sciences" (Goodwin, et al., 2008, p. 328). We prefer to fold reactivity to measurement into the ecological validity concept as it will similarly influence how close the experimental context is to the real-world situation of interest.

Table 1: Definitions of the dimensions of the Adapted ECOVAL Framework.

		Definition	Low (Artificial)	Medium	High (Natural)
Environmental context	**Environmental signals**	Sensory input from the environment (sounds, smells, etc.)	No signals	Synthetized signals	Real signals (dust, noise, heat, pain, etc.)
	Objects	Physical objects in the environment (furniture, building, etc.)	No objects	Mock objects	Real objects
Computer system	**Test medium**	Physical device used to interact with the system	Paper	Mock device / different device	Intended device
	User interface	Software	Video / Storyboard	Prototype / Mock-up	Final interface
What is required from the participants	**Task**	Experimental task performed by the participant	Only verbalized	Mimicked and possibly verbalized	Real usage Real manipulation
	Behavior	Behavior of the participants during the experiment	Only verbalized	Mimicked and possibly verbalized	Real actions (moving, talking, inspecting, etc.)
	Reactivity to measurement	Impact of the measurement processes on the research subject's behavior	Participant cannot act naturally or is restrained by equipment	Participant is aware of being studied but this does not affect his/her behavior much	Participant is unaware or able to forget the he/she is being studied

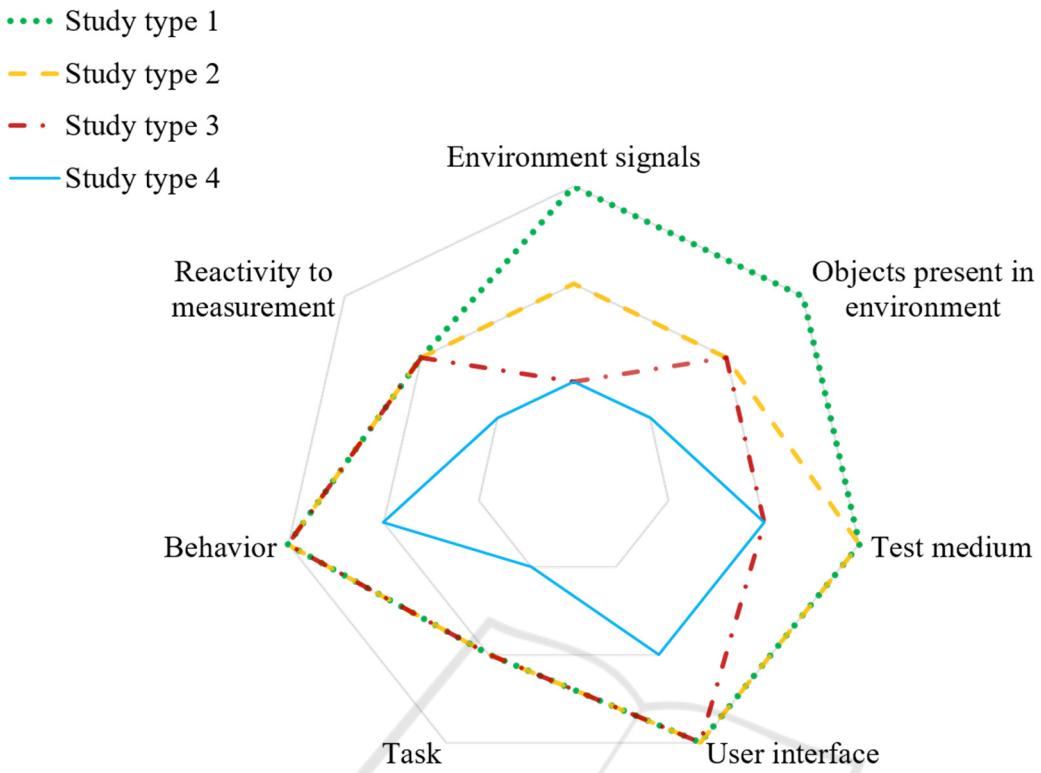

Figure 1: Adapted ECOVAL framework presenting 4 types of studies on 7 dimensions.

In Figure 1, you can find the modified Ecoval representation of 4 types of studies that were conducted at our user experience and physiological computing lab in recent years.

In the next section, we will present the different study types and how their level of ecological validity has led to data attrition and other issues. In addition, we will discuss lessons learned from these studies.

4 STUDY TYPES AND LESSONS LEARNED

To identify study types that can be grouped by their ecological validity, we selected a representative sample of 13 studies from the 128 that have been conducted at the Tech3Lab since its opening in 2013. Two senior researchers scored each of the studies on the 7 dimensions of the modified Ecoval framework. Building upon a grounded theory approach, we then identified the 4 types presented here and averaged the scores within a type to provide a score for the type itself which can be seen in image 1 (Glaser, 1992). Each type will be presented below along with some of the challenges associated with this level of ecological validity as well as some lessons learned.

4.1 Study Type 1: In the Wild

In study type 1, we find research projects that study very interactive technology that it is practically impossible to simulate in a controlled environment. In HCI research and in physiological computing research, the expression “in the wild” has often been used to connote research that is conducted outside of a laboratory and thus has a high level of ecological validity. For example, the study of a mobile game that requires geolocation such as Pokemon Go (Pourchon *et al.*, 2017). Physiological measures of these interactions need to be wireless and comfortable, allowing participants to move freely. In these experiments, we tend to use various combinations of eyetracking glasses (SensoMotoric Instruments, Teltow, Germany), portable cameras (Gopro Inc., San Mateo, CA, US), mobile dry EEG headsets (Cognionics Inc., San Diego, CA, US), and modular biosignal sensor kits (BITalino, Lisbon, Portugal). The main problems we encounter when conducting these studies are linked to the logistics involved in collecting minimally acceptable data quality. There are three categories of issues: data loss, poor data quality and synchronisation issues. Data loss is mostly linked to the limits of the equipment such as wireless transmission packet drop, maximum

recording durations, and equipment failure from overheating or water infiltration. Poor data quality can be linked to the participants more natural behaviour requiring adjustments such as positioning EDA sensors on the wrist rather than on the fingers or palm to allow subjects free use of their hands. Also, sweat can be a major issue, not only to obscure data from EEG and EDA, but also by reducing the adhesiveness of sensors.

In addition, the uncontrolled nature of the testing environment may affect the signals that are recorded. If we take for example pupil diameter, it is generally important to ensure that the measurement are reflective of an affective state rather than changes in lighting. That is not possible when it comes to sunshine. And as light can modify pupil diameter by 120% but affect can only change it by 20%, the latter will be drowned out and hard to distinguish (Laeng and Endestad, 2012). Finally, most wireless and portable equipment will be difficult to synchronize with others from different manufacturers. This leads to poorer precision in event markers which may limit the types of analyses performed, particularly with EEG.

Lessons learned:

- Do not assume anything! Even if the equipment is intended for your specific purpose, try it out ahead of time to record, export and analyse pilot data. This will ensure that the data quality is sufficient for the types of analyses you wish to perform.
- Schedule these projects at times when there are fewer chances of rain and high heat. This might reduce the ecological validity somewhat, but it will save your equipment from getting damaged.
- Public spaces may require you to obtain permits.
- Passers-by will ask you what you are doing, having an identified research team member will draw their attention away from the research participant.

4.2 Study Type 2: Simulated Wilderness

In study type 2, we include research projects that are conducted in the lab, in a simulated environment allowing the participants to physically interact with the technology. For example, a home cinema vibrokinetic system (Pauna *et al.*, 2018), a simulated virtual reality experience (Gardé *et al.*, 2018), or a wearable for a labourer in a simulated work environment (Passalacqua, Nacke and Leeger, 2018). As these experiments are conducted in the lab, the variety of equipment available is increased as is the control over ambient temperature, humidity and lighting. Many companies offer wireless data transmission for their equipment, but the receiver needs to be nearby, which makes allowing participants to roam freely in a large environment very difficult. When the subject is in a more confined space in the lab, this is much simpler.

Limits remain when it comes to eyetracking and facial emotion recognition which require a continuous and direct line of sight to the participants face. Eyetracking glasses can compensate for this somewhat, but are difficult to synchronize with other equipment and limit the scope of possible analysis. Other signal artefacts can come from allowing the research participants to move freely such as neck muscle strain which can be a problem in EEG, care should be taken when designing experiments where the subject has to bend over or turn their head regularly. Similarly, any muscle activity will lead to increased heart rate, even simply standing.

Lessons learned:

- Even with a state of the art EEG system with preamplified electrodes, much care has to be given to stabilizing the equipment on the participant. A movement of the wires of the EEG can lead to movement of the electrode itself causing significant artefacts.
- When trying to log participant activity as posthoc event markers, more cameras are preferable. It is very frustrating to realize when processing videos that a certain camera angle led to participants' actions being hidden by furniture or by participants bending over. However, more cameras mean heavier file weights and an increased need for disk space. When possible, decreasing the framerate of the recordings can help in this regard.

4.3 Study Type 3: Laboratory User Testing

Study type 3 includes more traditional user testing which is conducted while seated at a desk and requiring from the subject some fairly natural behaviour, such as navigating an online grocery store

and purchasing a list of items for a recipe they were given (Desrochers *et al.*, 2015). The level of ecological validity remains fairly high in this type of experiment as a person would probably be completing these tasks at home on a computer, while seated and the online shop is a real or close to real prototype. The behaviour and task are quite natural. The context itself is less so, as the person had to come to the laboratory and imagine themselves cooking this recipe in the future, all the while knowing they probably never will. To increase the ecological validity of this experiment would require going to the person's home and this increases the cost and complexity of acquiring physiological data drastically.

In terms of quantifying the ecological validity of this type of study, the main difference with other simulated environment type studies lies with the research question and with the physicality required in the interaction. The studies suited for this type are those exploring behaviours which would normally not require the participant to move around. This makes it much easier to capture high quality physiological data. Almost all physiological data types can be captured easily with this setup as the subject moves very little. There is no need for wireless equipment or particular camera angles. The limits for this type of study can come from the posture of the at home user that is not represented in the design. Think of an online shopping experience, while we could think that a neutral office type environment in a lab could be very representative of someone shopping online at their work computer, it is less representative of the type of shopping someone can do Friday night at home lounging on the couch with the TV in the background, a cell phone in hand and the computer precariously positioned on one knee. The user's posture itself will have an impact on the validity of the results, but also the emotional and attentional influences of the context. This would refer to the environmental signals and objects dimensions of the ECOVAL framework.

Lessons learned:

- Be conscious of the actual environmental context of the end user. That may even be an additional variable of interest for a study, where one portion of the study is conducted at a desk and another portion is conducted in a simulated living room. The study context can then be considered as a variable and its effect can be measured to see if participants react differently in the living room compared to seated at the desk.

4.4 Study Type 4: Isolated Cognitive Process

Study type 4 refers to classical cognitive psychology paradigms where the goal is to isolate the cognitive processes responsible for certain perceptions and behaviours. Experimental control is at its peak in this type of study. Results from these studies are more conclusive, but less generalizable. This level of control is indispensable when trying to understand specific cognitive processes. We have found this type of study useful as it allows researchers to better understand specific neural processes that underlie behaviour of interest using highly controlled experimental designs. In turn, this allows us to build a more solid scientific base on which to build more naturalistic studies. For example, a study using Transcranial direct-current stimulation to evaluate the contributions of a specific brain region in users' acceptance and trust of technology (Dumont *et al.*, 2014).

Lessons learned:

- This can be a good first step when trying to validate ground truth for a new method or tool as you can induce a given state in a very isolated manner (using a validated task to induce low and high levels of cognitive load).
- The lack of ecological validity in this type of study will need to be justified when submitting for publication. Researchers should not attempt to generalize their findings to "real-world" applications, but rather explain the benefits gained by this level of experimental control. They should also recommend how future research could extend their findings in more naturalistic settings.

5 RECOMMENDATIONS

First, projects in the wild, especially those that are outside, will generally require ad hoc modifications or adaptations to the equipment to allow for a personalized setup. Researchers from our team have performed these tasks: sewn a support pouch for an equipment, designed and 3D printed a case for an equipment, weather proofed said case only to realize in the end that it was simply not possible, thus this experiment could not move forward if rain was in the forecast, programmed a specific software to allow the synchronisation of two equipment, built a security

structure around a treadmill out of wood and pipes, and many other similar activities not generally expected from an academic outside of engineering.

As for devices, not so long ago, motion was among the main sources of artefact in signal recording (Healey, 2009). While some instruments have improved, motion remains a big problem for recording when you do not wish to tell the participant to refrain from certain movement. Solutions appear to be multiple. For signals with a small signal-to-noise ratio, such as EEG, preamplified electrodes can reduce some noise acquisition at the source. Also, specially developed algorithms for signal processing can be a great help (Bigdely-Shamlo *et al.*, 2015). To further the use of these signal processing algorithm we encourage a stronger dialog between method developers (engineers, statisticians, data analysts) and users of these methods (psychologists, applied neuroscience researchers, etc.) In addition, as suggested by van den Broek (2009), reducing the intrusiveness of sensors will improve ecological validity in the Reactivity to measurement dimension.

Secondly, a major limit to conducting studies in situ is the presence of uncontrolled elements that may threaten the safety of study participants. In some instances, this can be overcome through the use of confederates that are mindful of the person's safety or, in more complex circumstances, the use of simulators and virtual reality will come in handy. A different barrier then appears as virtual environments can be costly to program and implement properly.

Thirdly, increased ecological validity will often lead to increased data loss. Plan for this both in your time, in your participant recruitment and in your writing. You should explain the reasons behind a large data loss and how it was a worthy trade-off for the increased ecological validity.

Similarly, as a reviewer, data loss is not always synonymous with bad methods. Data loss should certainly be explained, but when properly justified by this trade-off, it should not be a reason to reject a paper. In addition, expensive methods are often associated with smaller sample sizes. Statistical analyses should be adjusted for this.

Finally, a bit of "food for thought". As research is, overall, a collective endeavour, one may wish to be careful that expensive and labour intensive in-the-wild research does not become the norm for research questions that do not require it. If every major publication on a given topic employs these methods, it will become expected, which may prevent younger researchers and less fortunate teams from pursuing these avenues.

6 CALL FOR RESEARCH

To continue to improve how ecological validity is optimized in physiological computing, many elements still need to be researched.

First, while it may seem that the highest level of ecological validity is often better for a project, a paper by Kjeldskov in 2004 and a follow up 10 years later showed that in some respects, in-lab simulated environments can be more conclusive than going in situ (Kjeldskov *et al.*, 2004; Kjeldskov and Skov, 2014). While these conclusions were drawn for usability studies, their observations on increased costs and man hours for outside-the-lab research are valid for all fields of research. To our knowledge, no such research comparing the complexities and costs of naturalistic vs. simulated environments has been conducted in physiological computing and this is certainly an interesting gap in the literature that needs filling.

Second, a common problem with very ecologically valid research is that by allowing participant to act naturally, they often do not repeat the same actions multiple times. Without repetition, there is uncertainty as to whether the responses are good representations of this stimulus in reality. Developing statistical methods able to extrapolate reactions based on each individual repetition instead of using aggregated measurements would be a great contribution to the field. In any case, aggregating a measure is less efficient than preserving all the individual measures and jointly analysing them through a longitudinal regression approach (Fitzmaurice, Laird and Ware, 2012), as long as we take into account the potential correlation between all the repeated measures. Also as stated by (Makeig *et al.*, 2009): "From a mathematical point of view, the basic problem is that complex functional relationships between two high-dimensional and highly variable signals (EEG and behaviour for example) cannot be well characterized by first reducing each signal to a few average measures and then comparing them. Rather, what is needed is a new and quite different approach incorporating better recording and modelling of relationships between high-density EEG and more natural and higher-fidelity behavioural recordings." (Makeig *et al.*, 2009, p. 4)

7 CONCLUSION

After a few years, of conducting as ecologically valid research as possible, we have come to see the benefits and challenges that accompany this type of research in physiological computing. And while it can sometimes seem like it is not worth it, in the end, a major reason to endeavour for a high level of ecological validity is the hope that a more authentic experience for the user will lead us closer to emotional ground truth, a famously elusive aspect of physiological computing (van den Broek, 2012).

That being said, the research questions and the theory underpinning a given research project should be the key factor in determining which dimensions of ecological validity are more of a priority. As such, papers should not be judged simply by whether or not they have strong ecological validity but rather as whether or not they have the appropriate ecological validity given the phenomenon studied.

As the technology keeps evolving and providing us with better research tools, we hope our advice can help other researchers design better studies and further the field of physiological computing.

REFERENCES

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M. and Robbins, K. A. (2015) 'The PREP pipeline: standardized preprocessing for large-scale EEG analysis', *Frontiers in Neuroinformatics*, 9(June), pp. 1–20. doi: 10.3389/fninf.2015.00016.

van den Broek, E. L. (2012) 'On the bodily expressions of emotion, be aware : More than a century of research! A commentary on Affective Computing by Kristina Hook', in *The Interaction-Design.org Encyclopedia of Human-Computer Interaction*. Aarhus C., Denmark: The Interaction-Design.org foundation. Available at: http://www.interaction-design.org/encyclopedia/affectivecomputing.html#egon+l.+van+den+broek.

van den Broek, E. L., Janssen, J. H. and Westerink, J. H. D. M. (2009) 'Guidelines for Affective Signal Processing (ASP): From lab to life', in *Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009*. IEEE, pp. 1–6. doi: 10.1109/ACII.2009.5349492.

Brofenbrenner, U. (1977) 'Toward an Experimental Ecology of Human Development', *American Psychologist*, 32(7), pp. 513–531. doi: 10.1037/0003-066X.32.7.513.

Brunswick, E. (1949) 'Ecological validity of potential cues and their utilization in perception', in *Systematic and Representative Design of Psychological Experiments*. Berkeley, USA: University of California Press, pp. 46–50.

Desrochers, C., Leger, P.-M., Senecal, S., Page, S.-A. and Mirhoseini, S. (2015) 'The Influence of Product Type, Mathematical Complexity, and Visual Attention on the Attitude toward the Website: The Case of Online Grocery Shopping', in *Proceedings of the Fourteenth Annual Workshop on HCI Research in MIS*. Fort Worth, Tx, US.

Dumont, L., Larochelle-Brunet, F., … H. T.-P. of the and 2014, U. (2014) 'Using transcranial direct current stimulation (tDCS) to assess the role of the dorsolateral prefrontal cortex in technology acceptance decisions: A pilot study', in *Proceedings of the Gmunden Retreat on NeuroIS*.

Fitzmaurice, G., Laird, N. and Ware, J. (2012) *Applied longitudinal analysis*.

Gardé, A., Léger, P.-M., Sénécal, S., Fredette, M., Labonté-Lemoyne, E., Courtemanche, F. and Ménard, J.-F. (2018) 'The Effects of a Vibro-Kinetic Multi-Sensory Experience in Passive Seated Vehicular Movement in a Virtual Reality Context', in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. New York, New York, USA: ACM Press, pp. 1–6. doi: 10.1145/3170427.3188638.

Glaser, B. (1992) *Basics of grounded theory analysis*. Mill Valley, CA USA: Sociology Press.

Goodwin, M. S., Velicer, W. F. and Intille, S. S. (2008) 'Telemetric monitoring in the behavior sciences', *Behavior Research Methods*, 40(1), pp. 328–341. doi: 10.3758/BRM.40.1.328.

Healey, J. A. (2009) 'Affect detection in the real world: Recording and processing physiological signals', *Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009*. doi: 10.1109/ACII.2009.5349496.

Kieffer, S. (2017) 'ECOVAL: Ecological Validity of Cues and Representative Design in User Experience Evaluations', *Transactions on Human-Computer Interaction*, 9(2), pp. 149–172.

Kieffer, S., Sangiorgi, U. B. and Vanderdonckt, J. (2015) 'ECOVAL: A framework for increasing the ecological validity in usability testing', *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2015–March(4), pp. 452–461. doi: 10.1109/HICSS.2015.61.

Kjeldskov, J. and Skov, M. B. (2014) 'Was it worth the hassle?', *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14*, pp. 43–52. doi: 10.1145/2628363.2628398.

Kjeldskov, J., Skov, M. B., Als, B. S. and Høegh, R. T. (2004) 'Is It Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field', in. Springer, Berlin, Heidelberg, pp. 61–73. doi: 10.1007/978-3-540-28637-0_6.

Laeng, B. and Endestad, T. (2012) 'Bright illusions reduce the eye's pupil', *Proceedings of the National Academy of Sciences*, 109(6), pp. 2162–2167. doi: 10.1073/pnas.1118298109.

Makeig, S., Gramann, K., Jung, T.-P., Sejnowski, T. J. and Poizner, H. (2009) 'Linking brain, mind and behavior.', *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*. Elsevier B.V., 73(2), pp. 95–100. doi: 10.1016/j.ijpsycho.2008.11.008.

Passalacqua, M., Nacke, L. E. and Leeger, P.-M. (2018) 'Gamification as a Catalyst for Collaboration', in *Bridging HCI*.

Pauna, H., Léger, P.-M., Sénécal, S., Fredette, M., Courtemanche, F., Chen, S.-L., Labonté-Lemoyne, É. and Ménard, J.-F. (2018) 'The Psychophysiological Effect of a Vibro-Kinetic Movie Experience: The Case of the D-BOX Movie Seat', in *Information Systems and Neuroscience*, pp. 1–7. doi: 10.1007/978-3-319-67431-5_1.

Pourchon, R., Léger, P.-M., Labonte-Lemoyne, E., Sénécal, S., Bellavance, F., Fredette, M. and Courtemanche, F. (2017) *Is augmented reality leading to more risky behaviors? An experiment with Pokémon Go*, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. doi: 10.1007/978-3-319-58481-2_27.

Schmuckler, M. A. (2001) 'What is Ecological Validity? A Dimensional Analysis', *Infancy*, 2(4), pp. 419–436. doi: 10.1207/S15327078IN0204_02.