

Proposing a novel approach for testing complex medical task trainer prototypes

Oscar Lilleløyken¹, Daniel Nygård Ege¹, Marius Auflem¹, Martin Steinert¹

¹Norwegian University of Science and Technology

oscarli@stud.ntnu.no

1 Introduction

In the early stages of engineering design, efforts are put into mitigating risk and preventing costly rework later in the design process. This early stage is often referred to as the Fuzzy Front End and is described by Kim and Wilemon (2002), as the period from when a product opportunity is first made apparent until it is deemed ready for further development. As this stage is early, unmatured, and highly ambiguous, risk mitigation concerns exploring multiple designs, by not being bound to one solution that might not work out (Sobek II et al., 1999). Prototypes are in this context important tools, not only to evaluate design ideas, but to map out and explore a solution space (Auflem et al., 2019; Leifer & Steinert, 2011). Limiting time and resources invested, multiple prototypes can be created while still informing designers about which concepts perform well, and which concepts are prone to fail (Leifer & Steinert, 2011). Such prototype-driven development methods show the importance of both divergent and convergent thinking (Eris, O., 2003). Divergent thinking enables designers to diverge from the concrete facts of the problem and generate multiple solutions that could help solve it. In the divergent space generative design questions are used to help idea generation. In the convergent space the aim is to converge on a fact-based foundation from which a solution can be identified.

While various stages of product design all face uncertain elements, this paper aims to address the lack of knowledge and insights needed to inform further development of concepts generated in the Fuzzy Front End. More specifically, it concerns user-centred design challenges within the field of medical training equipment, how they may be tackled using structured prototyping methods and how proposed solutions may be validated through user interactions.

1.1 Wayfaring

Wayfaring is a design framework encouraging freedom to explore and enlighten both problem and solution space in parallel as described by Steinert & Leifer (2012). Wayfaring suggests that iterative cycles of designing, building and testing of prototypes should be used to address uncertainty of facing new product opportunities. In this context, prototypes are used as learning probes to uncover and explore a design space and test concepts (Gerstenberg et al., 2015). Each probe investigates one or more ideas that in turn inform and steer further development. This approach allows flexibility, as well as the opportunity of finding a novel idea and/or uncovering a unique product potential in the early stages of design.

Two important concepts in this phase of development are known- and unknown- unknowns. While the known unknowns can be revealed through information-gathering such as consulting with experts, the unknown unknowns are hidden and cannot be revealed through inductive studies (Ramasesh & Browning, 2014). The unknown unknowns could, however, be uncovered by extensive prototyping as diverging within a broad solution space one is prone to encounter new challenges not yet on the radar. These unknown unknowns are important to uncover as they can have major unforeseen implications for a project. By Wayfaring, an understanding of product requirements may develop in a dynamic and emergent manner, and unknown unknowns be revealed as the project progresses (Kriesi et al., 2016).

1.2 Mixed method research approach

Mixed method research combines both quantitative and qualitative data, with the assumption that combining methodologies with different strengths and limitations can uncover more information from each test subject (Abowitz Deborah A. & Toole T. Michael, 2010). The reliability of the results is also increased, as it enables triangulation of the obtained data. Abowitz and Toole (2010) defines triangulation, in the context of research, as “using multiple research methods or measures to test the same hypothesis or finding”, exemplified by using both open-ended and closed-ended questions in a survey to measure the attributes in question. Dybvik (2018) describes the difference between quantitative and qualitative research;

- Qualitative research collects data from individual test subjects, and the subject’s personal experience interacting with the experiment or concerning the research issue at hand.
- Quantitative research collects empirical, numerical data from experiments or test subjects. An important difference is that data is collected, not by the researchers themselves, but through a research tool, e.g. questionnaires and sensors.

As the strengths in quantitative and qualitative research compensate for the weaknesses of the other, this combination is valuable (Abowitz and Toole, 2010). Though mixed method research can provide more data with higher reliability, Dybvik (2018) calls attention to the possible drawbacks of this approach compared to a single method. Most important is the added cost and time required to carry out the research, as well as the need for researchers with knowledge of different research methods.

1.3 Aim and scope

The scope of this paper is to explore and structure a way of testing conceptual prototypes of medical task trainers by using a mixed method approach, in order to guide further development. By presenting a case project where a medical task trainer has been developed, the use of this novel approach is exemplified. The aim of this paper is to discuss how to test conceptual prototypes to assess multiple functionalities and features in order to ensure that the required ecological validity of a medical task trainer is met. The paper will also show how the same approach can help developers make informed decisions for further design iterations when dealing with complex user-interactions, use-cases or other independent uncertainties. Furthermore, the implications for a novel approach to testing leveraging a mixed method research is discussed.

2 A novel approach to guide development by testing prototypes using a mixed method approach

User-testing and testing users' interaction with prototypes is complicated, as limiting sources of error and elements affecting users' feedback is challenging. During prototype testing, it is often necessary to interview and ask specific questions to obtain useful data for further development. In this context attention should be brought to the features that are being tested and whether the designers are collecting the necessary information. For example, will asking broad, open-ended, questions lead to new questions, rather than answer the ones crucial for further development? Houde & Hill (1997) emphasize the intent for each prototype, and that designers should be aware and reflective of what each prototype is testing. This paper argues that the same mindset should be applied when designing a *test* of a prototype - it should be tailored to the prototype at hand, and open design-questions that need answering. The novel approach presented in this paper elicits user feedback from testing by combining a Likert-scale questionnaire, sensor data and a post-test interview.

For many projects, a sufficient test of a conceptual prototype could be achieved by simplifying the prototype to only test individual features or functionalities at one time. Designers also have the option of simplifying and focusing prototype tests by using an approximated environment. As described by Vestad & Steinert (2019) prototyping the test environments themselves, and conducting proxy tests, can be useful to gain meaningful insights and understand how the prototype performs. However, in many cases considerable simplifications are not viable. Oversimplification may degrade the realism of a trial to a level where the observations made are no longer transferable to the actual situation the designer is interested in. This is especially relevant when prototyping products involved in complex real-world interactions and scenarios. The development of medical training equipment is one example. These products often require the simulation of human anatomy, and the interaction of the user with that simulation. In order to assess the perceived realism of anatomical features, a certain degree of fidelity is required for the test to be sufficient. If elements which affect the prototypes' physical attributes are neglected the entirety of the prototypes' realism is reduced. If the prototype cannot be made simpler, the test must be made more comprehensive.

In a novel approach to testing, high numbers of unconnected and uncertain elements are addressed by combining subjective and objective observations. Combining quantitative and qualitative data collected from the user is suggested to maximize usable information collected from each test subject. As the strengths in quantitative and qualitative research compensate for the weaknesses of the other, combining them can give a better chance of obtaining useful information (Abowitz & Toole, 2010). Quantitative data is collected from sensors in the prototype, and subjective opinions are collected and quantified through a post-test questionnaire. The questionnaire includes a Likert scale (Allen & Seaman, 2007) evaluation of prototype features. A Likert scale is a rating scale where the test subject rate in what degree it agrees with a statement regarding the product. Thus, whether the individual functionalities uncovered during development are present in the final conceptual prototype, can be evaluated. By using the Likert scale questionnaire, the subjective opinions of all test participants are also combined and presented as quantifiable data. The conceptual prototypes should also include sensors, so that objective data from the interaction between prototype and user can be collected. Qualitative data can also be collected as opinions from the users, either in written form or through interviews, or as observations made by the developers during testing. Thus, in-depth

explanation of why certain features was ranked more poorly than others can be collected. An overview of the new approach is shown in Fig. 1.

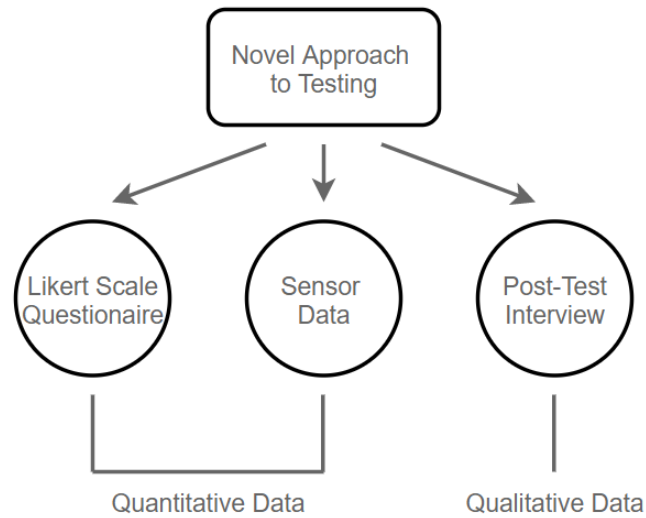


Fig.1. Illustration of a proposed novel approach for testing complex prototypes.

3 Case project- developing an abdominal examination task trainer

The case project concerns a product development project carried out in 2019 in Trondheim, Norway, by the authors to develop a tool to assist in abdominal palpation training. This project was initiated at TrollLABS, a product development laboratory at MTP, NTNU, in collaboration with a major European manufacturer of medical training equipment, as limited aids were available in medical education for the training using abdominal palpation to diagnose patients. This situation required users to train on human markers. While human markers enable high realism and close to real- world patient interactions; simulating diseases, pains, and altered tactile response is not possible. This project therefore wanted to explore the potentials of creating a "simulated patient", i.e., an abdominal task trainer allowing users to train their abdominal palpation diagnostic skills. One prototype task trainer developed over the course of this project is depicted in Fig. 2.



Fig 2. Conceptual prototype of an abdominal palpation task trainer.

This prototype had the overarching goal of fulfilling training goals from medical curricula, and obtain ecological validity, i.e. the required level of realism for users to benefit. Hence, multiple functions and design questions were attempted to be answered in parallel and represented in one conceptual model. To facilitate freedom to explore design space and work iteratively, the Wayfaring approach was chosen by the designers for this project. The workflow is presented in Fig. 3. The first step was to uncover what functionality and physical attributes an abdominal examination task trainer requires to meet the curriculum training goals. Through interviews and user interactions with prototypes, requirements were uncovered, and included both visual and tactile realism as well as ways of determining where pain occurs and its intensity. Physical and visual realism are both highly subjective experiences and a threshold for when something becomes “realistic” is very hard to pin down. Various ways of addressing these ambiguous qualities were explored in parallel in order to examine as many solutions as possible. Prototypes were used to address the uncertainties in how to achieve requirements. Iterations of different solutions was done in parallel, until the best solution emerged. Further iteration was completed before all solutions were collected and combined in a conceptual prototype. The development process of this conceptual prototype is described in further detail by Ege et al. (in press).

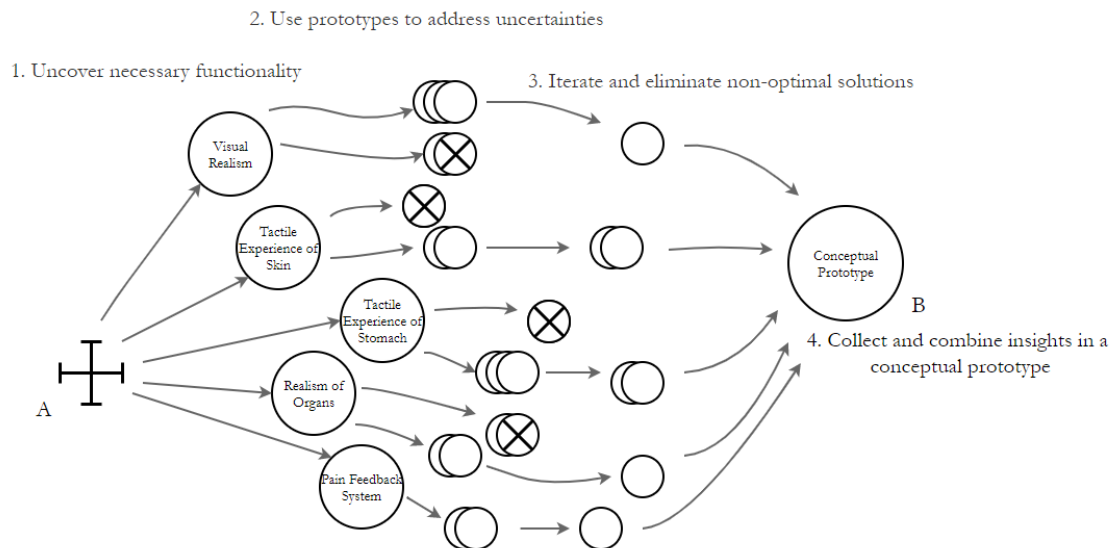


Fig 3 illustrates how parallel paths of prototypes with single functions are joined in a conceptual prototype.

The conceptual prototype developed consists of a soft stomach with an embedded sensor system. LEDs display pain intensity, while sensors embedded in the stomach measure applied pressure. The stomach is modelled to both look and feel anatomically correct and is integrated into a commercially available medical mannequin simulator. The prototype can simulate symptoms for 10 common abdominal diseases, including appendicitis and diverticulitis, as well as simulating inflamed organs.

The present case shows that in practice, the Wayfaring model which suggests development is like a winding path, can in fact be a collection of parallel paths taken in different directions of interest to answer design questions. Fig. 3 shows how parallel prototyping paths with single functions are joined in a conceptual prototype, combining insights and functionality. Throughout the project, prototypes were used to answer questions (Schrage, 2004). As insights and concepts are collected in one final prototype, the question of how well it incorporates all insights arise. Many of the features explored in the project are ambiguous, such as how something feels. Designers therefore relied on the opinions of expert users (in this context trained medical personnel), to prototype and recreate such attributes. Furthermore, a challenge that emerged throughout development was how to join the parallel prototyping paths. After a while, several loose ends had to be tied together, and compromises had to be made. E.g. in the presented project, incorporating the simulation of inflamed organs led to integration challenges because of a lack of space. This meant an optimal solution might have to be altered.

For the task trainer to accomplish its goal, i.e. facilitating the user to reach training goals, the right ecological validity must be present in the task trainer. Ecological validity is a measure of whether a test environment matches the real-world environment where the actual procedure is being performed (Kushniruk et al., 2013). It is important to note that only ecological validity in features important for creating effective training is in focus. Maximum ecological validity of all aspect of the task trainer is not expedient. Body heat, for example, was attempted included in the task trainer, but eventually abandoned as it added little to the learning outcome, while increasing complexity significantly. Each of the functionalities explored separately in development are meant to contribute to the ecological validity of the task trainer. To what degree the final conceptual prototype encompasses all the attributes is difficult to establish, as

it is based on the subjective input of few expert users. It is also difficult to predict how the features will affect each other. The test scenario is therefore complex and difficult to structure. Adding the subjective opinions of testers increases the difficulty of designing valid tests of such products.

This was the backdrop designers faced when aiming to test the developed conceptual prototype to establish its applicability to the intended educational setting, and if acceptable ecological validity was achieved. It was apparent that the designers needed a structured way for testing conceptual prototypes containing multiple unconnected unknowns. The approach presented in the next section can be generalized for how to test complex medical task trainer prototypes. Further, it can aid developers in answering questions to finish the Fuzzy Front End, by structuring both qualitative and quantitative data.

3.1 Novel approach applied for the case project

An approach to testing tailored to the conceptual prototype, as discussed previously, is presented below, and shows how the test approach can be used in practice. It simplifies test answers and helps structure results to address independent unknowns and inform further development phases. The novel approach can be applied to the abdominal examination trainer to verify that all requirements were adequately present in the conceptual prototype. The Likert scale questionnaire addressing functionality and physical attributes individually are shown in Fig. 4. Educational applicability is also covered in the questionnaire. A post-test interview allows participants to provide in-depth answers to support the questionnaire. This way follow-up questions could be asked regarding factors rated low in the questionnaire which could be valuable for development.

Using a scale from 1 (strongly disagree) to 5 (strongly agree), in what degree do you agree with the following statements concerning the palpation simulator

1 - Strongly disagree, 2 - Disagree, 3 - Neither agree nor disagree, 4 - Agree, 5 - Strongly agree

Statement		Rating				
		1	2	3	4	5
1	The simulator is visually realistic					
2	The skin feels realistic					
3	The stomach feels realistic					
4	The organs feel realistic					
5	The simulator is useful in training and education					
6	The simulator provided sufficient pain feedback					

Fig. 4. Abdominal task trainer post-test questionnaire.

The prototype includes four sensors located in each of the stomach's quadrants, shown in Fig. 5. The sensors track the relative applied pressure registered in each of the stomach quadrants over time. Analysing sensor data provides information on what areas of the stomach were

covered by the test subject, and whether the correct amount of force was used. This provides a better understanding of how well users perform and can be used as a measure of the training quality.

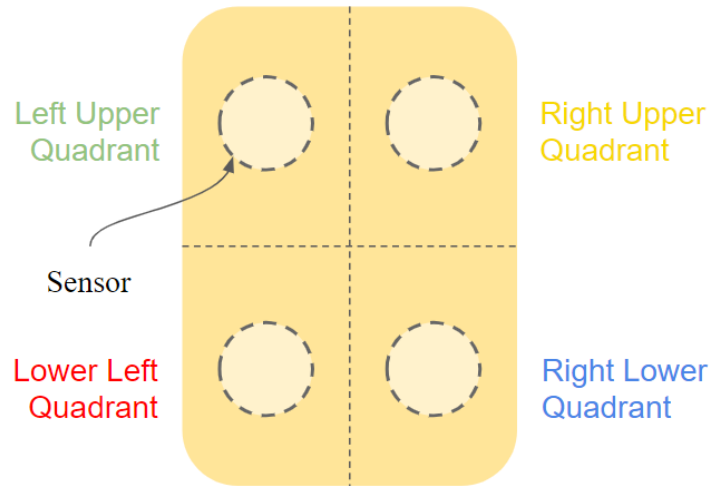


Fig. 5. Sensors placed in each quadrant of the stomach in the conceptual prototype.

Fig. 6. shows the data collected when a test subject was using the task trainer for different training cases. The X- and Y-axis show the time in milliseconds and relative pressure registered from each sensor respectively. The cyan line indicates what training case the user is working on, each increment shows the beginning of a new case. The cases and their correlating illness are shown below the X-axis.

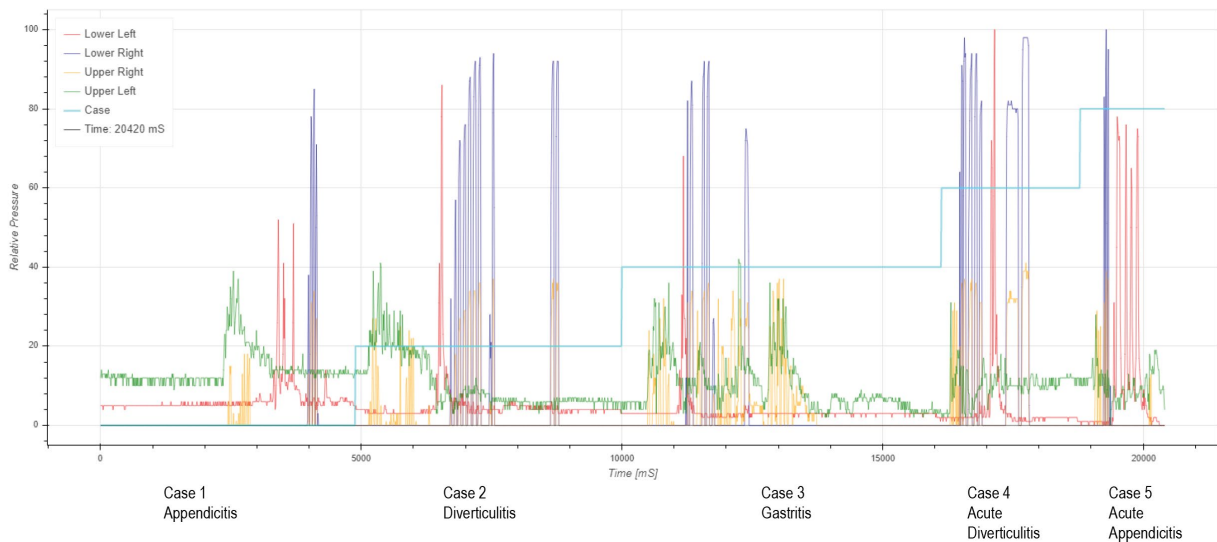


Fig. 6. The relative pressure registered by each sensor (Y-axis) over time (X-axis)

4 Discussion

In Wayfaring it is not self-evident how to perform effective tests as prototype resolution increases and as multiple questions are attempted answered within the same conceptual model.

However, performing these tests is still crucial as they might be solving known unknown challenges, validating known knowns or uncover unknown unknowns. It is therefore necessary to have enough tests in order to assure that previous iterations (and insights) are well represented and utilized in a concluding model. Prototypes exploring complex user-interactions and functionality, such as a medical task trainer, should not be tested in an overly simplified environment, test scenario or by leaving out critical product functionality. To confirm that the required ecological validity is present in a medical task trainer concept, a novel approach for testing of prototypes is proposed. Exemplified through a case project, the ambiguity of ecological validity has been addressed by utilizing a mixed method research approach and capturing both qualitative and quantitative data from user-testing.

In the case project, utilizing a mixed method research approach for prototype testing has been useful in order to inform future decisions and further development. The quantitative side of the applied method structures prototype feedback and illustrates how well a prototype performs based on results from all test subjects. Morse (2005), however, raises a concern of quantifying qualitative data by using forced-choice questionnaires. An alternative to giving a relative vote based on experience is to deploy multiple prototypes and hence use a ranked voting system. In terms of prototype resolution and complexity this could become costly – and time consuming, instead calling for a more comprehensive test. The qualitative side therefore helps explain and elaborate on the obtained results in the questionnaires. Further than merely investigating planned questions, unstructured interviews also allow following up on insight obtained during the actual test-procedure otherwise not captured.

Accommodating and leveraging sensor data is also important when performing complex prototype tests. In the context of this paper, sensor-integration in the prototype could enable both capturing and analysing how users interacted with the prototype. While these sensors were intended to generate feedback to users in training, collecting the sensor data could prove an important source of both insights and quantitative counterparts to the qualitative test results. This also calls for designers to be proactive when designing prototypes, and their corresponding tests, as sensors (redundant sensors) that might not seem valuable at the time could provide invaluable or supporting datapoints when analysing the results.

Traditionally in early-stage new product development the focus has been put on qualitative feedback collected from a limited number of key users. By making the data analysis more manageable through more structured data collection the novel approach proposed here enables collection of data from a greater number of users. An interesting aspect of collecting data in this fashion, is how to bring the qualitative and quantitative data from the test phase into the next phase of development. As the quantitative data can be used to prioritize the areas that need improvement, the qualitative feedback can shed light on why, and possibly how the improvement could be achieved. This allows features disrupting the ecological validity to be identified and solutions found based on feedback from the user.

Ambiguity and unknowns characterize the Fuzzy Front End, and whether the proposed test approach can mitigate some uncertainty from this development phase is worth consideration. By introducing objective and quantifiable data, unknowns and ambiguities can be identified and addressed in a structured and manageable way. Furthermore, it is interesting to see if this novel approach could aid projects in moving forward from the Fuzzy Front End. Eventually, when the needs have been elicited and functionality is identified, there is a need for measuring the prototype's performance and characteristics. Hopefully, the test approach can be used as a

decision-making tool for determining when the early stage development phase is over and more traditional project management and development tools can take over. Can a new approach to testing, tailored to the prototype or test in question, decide when this point has been reached? This question cannot yet be answered, but future test projects will hopefully shed more light on these issues.

A final consideration when applying the approach proposed is that by increasing prototype fidelity and capturing multiple design questions in a combined conceptual model, a more comprehensive test scenario would be required. Thus, the fidelity of the test should match the fidelity of the prototype in question. When prototyping in the Fuzzy Front End, extensive testing will grind the project to a halt, and time and effort spent testing complex prototypes is meaningless if not sufficient data is collected. Hence, whether an approach, such as described in this paper, is expedient or not is yet to be determined. The disadvantages described by Dybvik (2018), i.e. greater time consumption and increased cost, should therefore be carefully considered and compared to an alternative test regiment without mixed method testing. Pilot tests show promising results and have provided improvements to the proposed approach. It has however only been tested on a limited number of projects.

5 Conclusion

The main contribution of this paper is a tool to help developers of medical task trainers ensure necessary ecological validity for learning objectives. This tool is a novel approach to testing able to verify if all independently developed functionalities are present, when combined in a conceptual prototype. The case project presented in this paper exemplifies the need for such a method. The same approach has also been proposed as a way of informing decisions when faced with complex user interaction or ambiguous product requirements, by utilizing a mixed method research method. The advantages and implications of combining quantitative and qualitative data was also discussed. More research is necessary in order to verify that the proposed approach to testing provides adequate results. The community is therefore asked to use this novel approach in their development projects, to obtain more data.

6 Acknowledgements

This research is supported by the Research Council of Norway through its industrial PhD funding scheme, project number 290404.

7 References

1. Abowitz Deborah A., & Toole T. Michael. (2010). Mixed Method Research: Fundamental Issues of Design, Validity, and Reliability in Construction Research. *Journal of Construction Engineering and Management*, 136(1), 108–116.
2. Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality progress*, 40(7), 64-65.
3. Auflem, M., Erichsen, J. F., & Steinert, M. (2019). Exemplifying Prototype-Driven Development through Concepts for Medical Training Simulators. *Procedia CIRP*, 84, 572-578.

4. Dybvik, H. (2018). An exploration of mixed methods, researching business model phenomena-Investigating the business model phenomena through literature, in-depth case study, statistical testing and user experiments (Master's thesis, NTNU).
5. Ege, D.N., Lilleløyken, O., Auflem, M., & Steinert, M. (in press). Dealing with Ecological Validity and User Needs when Developing Simulation Based Training Equipment – Case Study of a Medical Palpation Task Trainer. *Procedia Cirp*.
6. Eris, O. (2003). Manifestation of divergent-convergent thinking in question asking and decision making processes of design teams: A performance dimension. In *Human Behaviour in Design* (pp. 142-153). Springer, Berlin, Heidelberg.
7. Gerstenberg, A., Sjöman, H., Reime, T., Abrahamsson, P., & Steinert, M. (2015). A Simultaneous, Multidisciplinary Development and Design Journey – Reflections on Prototyping. In K. Chorianopoulos, M. Divitini, J. Baalsrud Hauge, L. Jaccheri, & R. Malaka (Eds.), *Entertainment Computing—ICEC 2015* (pp. 409–416). Springer International Publishing.
8. Houde, S., & Hill, C. (1997). What do Prototypes Prototype? In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of Human-Computer Interaction (Second Edition)* (pp. 367–381). North-Holland.
9. Kim, J., & Wilemon, D. (2002). Focusing the fuzzy front-end in new product development. *R&D Management*, 32(4), 269–279.
10. Kriesi, C., Blindheim, J., Bjelland, Ø., & Steinert, M. (2016). Creating Dynamic Requirements through Iteratively Prototyping Critical Functionalities. *Procedia CIRP*, 50, 790–795.
11. Kushniruk, A., Nohr, C., Jensen, S., & Borycki, E. M. (2013). From Usability Testing to Clinical Simulations: Bringing Context into the Design and Evaluation of Usable and Safe Health Information Technologies: Contribution of the IMIA Human Factors Engineering for Healthcare Informatics Working Group. *Yearbook of Medical Informatics*, 22(01), 78–85.
12. Leifer, L., & Steinert, M. (2011). Dancing with Ambiguity: Causality Behavior, Design Thinking, and Triple-Loop-Learning. *Information-Knowledge-Systems Management*, 10, 151–173.
13. Morse, J. M. (2005). Evolving Trends in Qualitative Research: Advances in Mixed-Method Design. *Qualitative Health Research*, 15(5), 583–585.
14. Ramasesh, R. V., & Browning, T. R. (2014). A conceptual framework for tackling knowable unknown unknowns in project management. *Journal of Operations Management*, 32(4), 190–204.
15. Schrage, M. (2004). Never go to a client meeting without a prototype [software prototyping]. *IEEE Software*, 21(2), 42-45.
16. Sobek II, D. K., Ward, A. C., & Liker, J. K. (1999). Toyota's principles of set-based concurrent engineering. *MIT Sloan Management Review*, 40(2), 67.
17. Steinert, M., & Leifer, L. (2012). 'Finding One's Way': Re-Discovering a Hunter-Gatherer Model based on Wayfaring. *International Journal of Engineering Education*, 28, 251–252.
18. Vestad, H., & Steinert, M. (2019). Creating your Own Tools: Prototyping Environments for Prototype Testing. *Procedia CIRP*, 84, 707–712.

