CASE STUDY

User involvement in early-stage design of medical training devices – case of a palpation task trainer prototype

Daniel Nygård Ege, M. Auflem, O. Lilleløkken and M. Steinert

Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT
Design of new medical training equipment is demanding as unavoidable complexity and ambiguity facing designers must be addressed in the early stages of the development. In this paper, a novel concept for abdominal palpation training is presented and used to exemplify challenges and approaches for designing new medical training equipment. Concluding the initial development of the palpation training concept, experienced medical personnel evaluated a conceptual prototype. A Likert scale questionnaire, a clinical assessment by participants, and recorded sensor data were used to evaluate prototype functionalities, perceived tactile- and visual realism, and usability of the concept in medical training. Obtained results are used to discuss insights for further development of the concept. Further, the paper discusses observations from user interactions and considerations regarding fidelity of medical training equipment prototypes. Moreover, it highlights the benefits of utilizing mixed-method research to identify areas of improvement for conceptual prototypes even before initiating industrial development efforts. By this, designers can ensure that user needs, product requirements, and sufficient fidelity are collectively captured in new medical training equipment in the early design stages.

ARTICLE HISTORY
Received 25 March 2021
Accepted 2 September 2021

KEYWORDS
Medical simulation; abdominal palpation; simulation-based training equipment; product development; user involvement; early-stage design

Background
In medical education, repetition and applying theoretical knowledge through skill training are essential for various procedures, including abdominal palpation (Cooper and Taqueti 2004). As it relies on tacit knowledge and recognizing distinct symptoms and tactile sensations, abdominal palpation is a fundamental skill that requires hands-on training (Aubin, Gagnon, and Morin...
In the procedure, practitioners use fingers and palms to localize areas with abnormal tactility (due to inflammation, anomalies, trapped fluids, and others) or areas causing pain to the patient. Abdominal palpation training (APT) is thus crucial for teaching primary diagnostics and for students to practice psychomotor skills and procedural algorithms in a safe and repeatable environment. This calls for solutions enabling realistic physical interactions, as the tactile sensations and techniques are hard to formalize and describe by theory alone.

APT is generally performed by fellow students acting as the patient to prevent risk and discomfort for actual patients, limiting the trainee’s objective feedback. Furthermore, such training does not capture the tactility of abdominal symptoms and corresponding illnesses. There are examples of research and development efforts attempting to provide APT equipment for education, proposing physical (Yakubo et al. 2008; Arita et al. 2019; ABSIM 2020; Laerdal 2021), virtual (Dinsmore et al. 1997), and hybrid patient simulation solutions (Ribeiro et al. 2016; Escobar-Castillejos et al. 2016; Ullrich and Kuhlen 2012). However, these solutions are generally expensive, non-continuously available, or lacking critical functionality such as objective performance feedback or tactile/physical feedback. Another potential issue for the current training solutions is the finite set of study cases to practice on, and not being able to tailor these to the various disciplinary and contextual norms of APT.

Given the current APT challenges, a leading provider of medical training equipment proposed a new development project to explore and conceptualize a solution for simulation-based APT. This project developed a concept through iterative prototyping and evaluated it using a multi-faceted test framework for the user-centred design of medical training equipment. Development and testing revealed the need for an easy-to-use solution that could provide users with objective feedback, enabling a large (and expandable) set of study cases and facilitate realistic tactile interactions. The resulting conceptual prototype is a low-cost implementation of readily available hardware, software, and materials that could be developed further as an open hardware solution in the future.

As the design of equipment for medical training can impact the performance of healthcare professionals and thus patient safety (Leape et al. 1991; Lawton et al. 2012; Zhang et al. 2003), it is essential to ensure sufficient clinical realism through user involvement in the development process (Money et al. 2011; Blandford, Furniss, and Vincent 2014). However, sufficient realism can be difficult to manage, given a large chasm between designers and healthcare professionals (Zenios et al. 2009). Although similarities between healthcare and design exist (Rowe, Knox, and Harvey 2020), generally, differences in methodology and procedures are perceived as challenging for designers (Groeneveld et al. 2018).
This paper presents how structured prototype testing with expert users can address the disciplinary gap between healthcare professionals and designers. Following the initial development of an APT concept, experienced medical personnel assessed a prototype and conducted a diagnosis for a set of different simulated conditions. The users were then requested to evaluate and provide feedback on the presented concept. This study aimed to answer whether the concept met the needs and requirements of healthcare professionals and if the prototype provided sufficient fidelity and clinical realism to be used in APT. With this, we provide an example of how structured user testing can evaluate the required functionalities and fidelity of prototypes during the development of new medical training equipment.

Development of prototype

In response to the highlighted challenges with APT, Ege et al. (2020) has developed a novel simulator concept to be used in medical education (Figure 1). This development project used an iterative and prototype-driven approach, accommodating the knowledge gap and ambiguity facing design teams in the early stages of product development (Gerstenberg et al. 2015; Motavalli and Nestel 2016). Multiple low-resolution prototypes were designed, built, and tested to elicit and explore functional requirements and necessary fidelity (Ege et al. 2020).

Figure 1. The APT simulator developed by Ege et al. (2020) and assessed by expert users in this paper.
The simulator consists of a soft stomach with an embedded sensor system. Three layers of different silicone rubbers make up the stomach and skin. It encompasses a force sensor in each quadrant of the stomach, measuring both forces applied when palpated and their location. Sensor readings were correlated to pain levels by having expert users interact with early prototypes and suggest how painful the pressure they applied would feel. The sensor’s variable voltage is measured and mapped into ten pain levels, ranging from low amounts of pain to high. An array of ten LEDs light up to visualize pain intensity during an examination of the prototype. Simulated palpable organs, including the liver and gall-bladder, can be inflated to indicate inflammation, giving the task trainer the possibility of simulating various diseases and symptoms. A total of 10 common abdominal diseases are currently implemented, including appendicitis, diverticulitis, and gastritis, and could easily be increased or changed based on the need of educators. The task trainer can simulate both mild and acute symptoms of appendicitis and diverticulitis. The intended disease corresponds to typical symptoms such as pain intensity and location and signs of rebound tenderness and enlarged organs. A computer logs sensor data allowing for post-test analysis of trainees. The task trainer is integrated into a modified commercially available medical mannequin, modelled to look and feel realistic with anatomical landmarks such as ribs and hip bones.

Medical professionals were consulted in the earliest phases of design to elicit requirements for a new training device. Figure 2 illustrates how prototyping activities guided development, building on previous findings and requirements. In addition, it shows how prototypes addressed uncertainty by building multiple solutions and testing them to pick the best one before combining all sub-systems in a final prototype.

While quantifying attributes such as tactility and the visual resemblance is challenging, designers captured an appropriate resemblance in a final
prototype by building and testing multiple rapid and low-cost prototypes qualitatively. An initial, rough prototype consisting of foams, waterfilled condoms, and a skin-coloured PVC sheet (Figure 3(a)) was built as an initial idea of what an APT simulator could be. As well as being a valuable discussion point when meeting experts, this prototype yielded important insights challenging to capture using quantitative measures, such as the palpation technique and tactile realism. Likewise, solutions to simulate organs changing tactility due to disease were also prototyped using simple, readily available materials. For example, based on a doctor describing an inflamed gallbladder feeling like a soft plum, several prototypes were made to replicate what the designers believed it should feel like before experts confirmed or invalidated its tactile realism. A prototype aiming to simulate an inflamed gallbladder is showed in Figure 3(b), where a balloon filled with water replicates the tactility the expert described.

Iterative prototyping and continuous user interactions, focussing on hands-on testing and interviews, guided development towards the conceptual prototype presented in this paper. Early designs with increasing complexity and realism are showed in Figure 4, where more comprehensive sensor systems and more realistic anatomical landmarks were included as development progressed.

The development of this task trainer prototype supports that before investing time and resources into building refined models/products, low-resolution prototypes should be rapidly developed and evaluated by users (Ege et al. 2020; Auflem, Erichsen, and Steinert 2019). This could aid both eliciting and solving concrete user needs and mitigate the risk of encountering severe obstacles later in the development (where it is more costly and time-consuming to address these) (Takeuchi and Nonaka 1986; Steinert and Leifer 2012; Thomke and Reinertsen 1998). In order to keep development agile and rapid, qualitative methods such as interviews and hands-on testing proved helpful in this project. User testing requires both reflective and intentional prototyping approaches as users are unaware of which aspects of the design idea the prototype represents (Houde and Hill 1997; Reay et al. 2017).

Figure 3. (a, b) Low-fidelity prototypes built early in the project.
More so, it is not evident which level of resolution (amount of detail) and fidelity (closeness to eventual design) a prototype requires to sufficiently convey a design idea and be deemed realistic enough to gain feedback. Especially when designing new medical training equipment, products mainly involved in complex real-world interactions, oversimplification may degrade the realism of a test to a level that is no longer valid (Lilleløkken et al. 2020). In early-stage development, qualitative feedback is often collected from a small sample of key users. However, increasingly complex prototypes call for increasingly comprehensive test procedures to accommodate them. Lilleløkken et al. (2020) suggest using a multi-faceted test framework utilizing quantitative methods strategically, as done in this paper.

Method

A mixed-method research approach generated three separate data sets: sensor data, questionnaire answers, and a clinical assessment performed by participants. Both qualitative and quantitative data were gathered to guide further development iterations. This research project was conducted over a 10-week period, where testing was performed with four participant groups of healthcare professionals, constituted by physicians and nurses, ranging from novice to experts.

Participants

Eighteen healthcare professionals participated in this study. Subjects were divided into four groups based on their profession and level of experience. G1 was a gastroenterologist with 40 years of experience. G2 consisted of newly educated physicians doing their first year of hospital training. G3 were general practitioners with an average of 8 years of experience. G4 consisted of experienced nurses, in administrative and educational positions, with an
average of 16.8 years of experience. Table 1 presents demographic information and how often participants perform abdominal examinations. Item D5, Expertise, describes the number of participants that perform abdominal palpation weekly from each participation group.

**Procedure**

Data were collected during four separate sessions, one for each participant group, in a similar environment, and a hospital setting. All examinations were done individually in a separate room from other participants. Participants were introduced to the APT simulator before how it works and how it indicates pain was demonstrated. It was then prepared to simulate different diseases. Because of different expertise and availability, the procedure varied between participation groups. G1 and G2 followed Scenario 1, while G3 and G4 followed Scenario 2, as illustrated in Figure 5.

In Scenario 1, participants from G1 and G2 were first presented with a list of 10 common abdominal diseases, shown in Figure 6(a). They were asked to examine the task trainer six times and propose a diagnosis from the list for each examination. Participants were allowed to use as much time as they wanted to gather findings and to determine diagnoses. The order in which diseases were simulated was the same for all participants.

In Scenario 2, participants from G3 and G4 were asked to examine the task trainer twice. For each examination, they proposed a diagnosis. Participants were allowed to use as much time as they wanted to gather findings and to determine diagnoses. The order of the simulated diseases was the same for all the participants in each test group.

Immediately after examinations, all participants answered a 5-point Likert scale questionnaire.

**Results**

**Clinical assessments**

After each examination, participants were asked to propose a diagnosis for the task trainer. Table 2 shows how many of the participants gave the correct diagnosis for each of the different diseases that were simulated.
The participants from G1 and G2 who followed the procedure in Scenario 1 could score six correct diagnoses. The participant from G1 scored six out of six correct diagnoses. From G2, three participants managed to provide five correct diagnoses, three provided four correct diagnoses, two provided three correct diagnoses, while the last participant managed to provide two correct diagnoses. The mean accuracy score was 4.1 out of six possible, with a standard deviation of 1.11.
A questionnaire determined how well participants agreed with six statements regarding the prototype’s physical, visual, and conceptual level. A 5-point Likert scale, where 1 = strongly disagree and 5 = strongly agree, was used to rate the level of agreement to each statement.

Table 3 summarizes questionnaire answers for all participant groups. Participants agreed the strongest with the question regarding usefulness in education, while they disagreed most with the realism in the feel of stomach and organs.

On average, G1 agreed most with the statements in the questionnaire (M = 4.17, SD = 0.98). G4 (M = 3.58, SD = 0.48) and G2 (M = 3.15, SD = 0.72) agreed less with the statements, while G3 (M = 3.0, SD = 0.84) agreed the least to the statement of all participating groups.

Collected sensor data

From a development context, sensor data was not intended for post-test analysis but rather to make the prototype function as intended. However, observed trends in the recorded sensor data show potential for uses exceeding the primary goals. Each sensor registers forces applied over time. Therefore, visualizing readings from all the sensors provides an overview of how the user performed the examination. In the visualization, the forces exerted on each of the sensors are visible, showing the palpation activity (either magnitude or location) in relative proximity to each of the sensors. Further timing, such as rhythm and time to trigger all sensors, can be observed in the raw data visualization. An example of the raw data gathered
from sensors during a test scenario where one disease was examined is depicted in Figure 7.

Cumulating the magnitude of force applied to each quadrant can visualize how evenly an examination is performed and whether the subject focuses equally on each quadrant. In Figure 8, the circles in each quadrant relate to the total force registered from each sensor over a simulated case. Similarly sized circles are positive, while variable sizes show uneven focus during examination. Figure 8(a) shows the plot of an expert user examining the task trainer for appendicitis. Figure 8(b) show a novice user’s examination.

An observation made during testing was that the amount of force applied to the task trainer when the pain was identified increased. Several participants seemed to try to max out the pain scale when detecting pain in a region, possibly to verify that the pain was indeed present.

Discussion

Interpretation of results

This study shows that the task trainer can be used for diagnosis practice and symptom recognition successfully. Results, therefore, indicate that it has the
functionality and attributes necessary for facilitating APT. However, it is not possible to ascertain the learning outcome of using the task trainer at this stage and thus its training potential. When comparing participants in the clinical assessment, it is clear that diagnosing appendicitis and diverticulitis yielded high success rates. This might be caused by the fact that these diseases have distinct symptoms different from other diseases. Regardless, the task trainer seems to replicate these distinct symptoms adequately. Also apparent in the obtained results is that high success rates can be seen for both acute and milder symptoms of appendicitis and diverticulitis, showing continuity in answers. This strengthens the hypothesis that simulated symptoms are adequately incorporated in the task trainer.

In the obtained questionnaire answers, participants neither strongly agree nor disagree with most statements. An interesting observation is that all groups disagree the most with the same statement (Q4). Similarly, they agree the most with the same statement (Q5). Participants not strongly agreeing with certain statements is not unexpected, given that the presented task trainer is a prototype and not a finished commercially available product. Nevertheless, the results are still of interest for further development, as they highlight parts of the concept with the most considerable improvement potential. For instance, improving tactile realism of enlarged organs is of interest as the users rated this functionality poorly. Further, participants agreeing to the usefulness of the concept in education is promising, suggesting further development. These insights illustrate why a Likert scale questionnaire was utilized. It proved a convenient way of capturing participants’ feedback on the current prototype’s physical, visual, and conceptual traits.

Limiting the amount of pain endured by a patient throughout an examination is essential in a clinical examination. By investigating the sensor data obtained during testing, how much force each participant has exerted can be determined, making it possible to detect whether the patient experienced excessive pain. Informing trainees of this during training with a task trainer can make them aware of how much force to apply during palpation. Secondly, the pattern of how trainees divide the abdomen and examine it quadrant by quadrant can also be derived from sensor data. This can be used to determine if a trainee methodically examines each quadrant, one at a time, to make sure the entire stomach is palpated. This way, he/she can be sure of where pain occurs and can determine a diagnose. The examination pattern also allows for comparison between experienced and inexperienced users, which could provide meaningful data for creating self-directed learning scenarios. The amount of time used on each scenario is also evident in sensor data, giving yet another basis for comparison.

How well quadrants have been covered by palpation can be derived by combining the total force applied to each quadrant. This can be used to
visualize uneven examination for trainees. It also allows for comparison with an expert and can highlight incorrect techniques. For example, Figure 8(a) shows how an expert pays equal attention to each quadrant. However, in Figure 8(b), a novice user is seen to focus more on some parts of the abdomen, possibly missing out on vital information to determine a correct diagnosis. These findings in the obtained data should be investigated further, as they could be valuable as objective performance measures to novice trainees and to give corrective feedback during and post APT scenarios.

**Comments on the prototype**

The prototype’s usefulness as a learning tool was rated highly across all participation groups, indicating its applicability in education. In addition, most users also agreed that the prototype’s ability to indicate pain was sufficient. Detection of pain has been described as the most prominent part of abdominal examinations, as the location of pain often is the most telling symptom for common abdominal illnesses. With the ratings of the other assessment points scoring around the middle of the 5-point scale, it is an indication that the users find the features offered by the prototype sufficient for practising the clinical skill for palpation.

As shown in the results, the force applied to the task trainer when the pain was identified increased. In a real-life scenario, the examiner would likely have gradually increased the force applied in painful regions, careful not to cause excessive pain to the patient. However, during testing of the task trainer, it recorded high pain levels in many cases even after the pain was located, as previously described. This behaviour is a sign of wrong technique and should be addressed by either a supervisor during training or by the task trainer itself. If used as a self-directed learning tool, corrective feedback should be given to trainees as avoidable pain to any actual patient should be mitigated. This phenomenon might be caused by a lack of immersion in the training scenario and lacking a more natural pain response. Test subjects were seen to be immersed going into the case. The physical appearance of the task trainer is realistic, as it is integrated into a full-size torso with accurate anatomical landmarks. Some participants were even concerned with warming up their hands before touching the ‘patient’ during an examination. However, the realism of the LED pain response significantly differs from the rest of the task trainer. This can have caused subjects to begin experiencing the case more as a game of ‘finding the right spot’ rather than a real scenario with a patient. The objective of the simulation can get lost and become overshadowed by the participants’ will to succeed in what they experience as a game (Frank 2012; Rieber and Noah 2008). Human pain response includes all senses, including audible responses, visual responses
like facial expression or body movement, and haptic responses from guarding, that is, muscle contraction in response to pain. Pain feedback from the task trainer is only addressing the visual aspect of the experience. Therefore, other sensory elements might need to be included in the future development to increase realism. This is in line with what (Stokes-Parish, Duvivier, and Jolly 2020) found, where a lack of immersion caused students to ‘do tasks for the sake of doing it’ during simulation, as opposed to performing as if the simulator was an actual patient.

**Comments on fidelity**

Simulation of clinical practices might be misrepresented as relatively simplistic (Motavalli and Nestel 2016), making it essential for designers of training equipment to acknowledge and consider the complexity and ambiguity of clinical simulation. Parts of this ambiguity and complexity can be attributed to finding the appropriate fidelity of a new medical simulator.

The fidelity of a simulator has been shown to affect trainee’s engagement in training, where more realistic simulators tend to increase engagement and a more immersive experience (Stokes-Parish, Duvivier, and Jolly 2020; Hagiwara et al. 2016). Immersion is essential in simulation as it can improve knowledge transfer to real-world scenarios (Dede 2009). However, an increase in fidelity does not necessarily lead to a higher learning outcome (Issenberg et al. 2005; Hamstra et al. 2014; Scerbo and Dawson 2007; Norman, Dore, and Grierson 2012). Although different methods for benchmarking simulator fidelity exists (Wilson et al. 2018), in the early development stages, it is not evident what the required fidelity for a simulator is. It can, however, as shown in this paper, be elicited through early-stage user testing.

The aspect of fidelity is somewhat difficult to comprehend in the early design stages. It concerns both the development stage a prototype is in and the intended fidelity of a finished product. The fact that a prototype is being tested instead of a finished product will influence results. This introduces a significant limitation in this stage of product development, as building high-fidelity prototypes would be too costly. Nevertheless, results in this study show that functionality and fidelity are within what users believe is sufficient without relying on a high-fidelity prototype. This paper advocate that designers should focus on finding what is realistic enough to obtain valid test results during the early development of medical training equipment, as opposed to being as realistic as possible. In this way, designers can avoid unnecessary complexity and expensive, time-consuming challenges that could be addressed at a later stage of development.

In this development project, prototypes were tested throughout by expert users, thus pinpointing what had to be improved for each prototype
iteration. In order to conclude early development, ensuring that functionality, appropriate fidelity, and fit within curricula are all sufficiently addressed in a concept, structured testing proved crucial, as it answered these questions.

**Implications on the further development**

The conceptual prototype presented in this paper has shown to be useful for tactile recognition and abdominal diagnosis exercise. It has been suggested that the increased fidelity in a task trainer does not necessarily result in better learning outcomes. Thus, future development will aim to create realistic training approaches and scenarios that can facilitate effective training. Moreover, the intent is to find ways to measure training outcomes to inform both trainees and educators.

To measure training outcome, a benchmark to compare participants against is needed. By comparing sensor data from participants with a gold standard examination recorded in expert user testing, trainees can be assessed. This can also allow for supplementing deficit-oriented corrective training approaches usually characterized with simulators, with an approach that focuses more on how good performance is produced, as described by (Dieckmann et al. 2017). Furthermore, sensor data could inform neural networks and algorithms, where pattern recognition can provide feedback and corrective suggestions. Location of touch, the areas covered by examination, peak forces, and timing is already possible to capture using the integrated sensors. This data could enable algorithms to locate not only simple pain points but also follow complete diagnostic scenarios.

Testing revealed that the simulation of enlarged organs was not satisfactorily captured in this prototype. Most participants did not even find the enlarged organs or realize that the simulator could emulate them. This exemplifies the importance of training before introducing a novel device, as discussed by Reid-McDermott et al. (2019), as trainees know what the simulator will test. It also means it might be necessary to exaggerate this functionality, that is, increase the size of organs to make them more distinguishable by trainees. This functionality could also be more dynamic where size and hardness could be altered to ensure a specific level of difficulty. Hence, training could be tailored to a specific scenario, curricula, or even individual skill levels.

**Limitations**

This study aimed to support further development of a conceptual prototype and not assert if trainees improve in abdominal examination using the prototype. Thus, it was considered redundant to correlate questionnaire data, sensor data, and clinical assessment of participants. This, for instance, implies
that it is unknown how the sensor data of successful participants compare to those who are less successful in diagnosing the task trainer. However, it is of great interest for future tests as it speaks to the prototype’s value as a learning tool. The ability to distinguish between skill levels is necessary to provide feedback to the user. Identifying areas of improvement can only be achieved by locating where a trainee’s input varies from a standard for successful examinations.

Data presented here is based on small and unbalanced groups, thus limiting its general validity. However, test data was primarily intended to enlighten further development of a prototype, and it was not considered necessary to increase the number of participants. As previously shown, the data gathered has proved helpful in pinpointing shortcomings of the current prototype and eliciting ways of going forward, thus fulfilling its intention. Due to the uncertainty caused by small and unbalanced groups, the findings should not necessarily be generalized to all other projects of medical training equipment. However, as this way of developing and testing medical task trainer prototypes proved helpful in this case project, it might be worth investigating further in other cases.

**Conclusion**

This paper has presented the results of structured testing of an APT simulation prototype with expert users. Participants performed an abdominal examination of the prototype and provided a diagnosis based on their findings. Three data set were collected from each participant; a clinical assessment, a Likert scale rating of the prototype’s features, and sensor data collected by the prototype. We found that participants correlated the simulated symptoms with high success rates to several medical diagnoses attempted recreated from the assessment. The continuity of these results further supports the concept as a valuable tool for the enabled practice of abdominal examination and to diagnose based on this procedure.

Further, subjective feedback from the questionnaire has highlighted areas of improvement for the conceptual prototype. The questionnaire results also support the concept as valuable for use in medical education. Obtained sensor data from the tests allow for technique and routine to be investigated in relation to the assessment results. It is discussed how this type of data could be leveraged for developing self-directed learning scenarios and objective feedback based on palpation data generated by experienced physicians. Using this data, insights concerning accumulated palpation force exerted (relating to pain), timing and coverage of the abdomen are observed. The prototypes’ functionality and implications, and considerations for further development of the concept are discussed based on findings from the test
results. By the mixed-method structured testing, insights and considerations for developing medical training equipment are also exemplified by this case project. Finally, we discuss why the findings in this project should not necessarily be generalized to all other projects of medical training equipment.

Disclosure statement
No potential conflict of interest was reported by the author(s).

Funding
The Research Council of Norway supports this research financially through its industrial PhD funding scheme, project numbers 290404 and 321386.

Notes on contributors

Daniel Nygård Ege works in Nasjonalparken Næringshage as an industrial PhD candidate through the Department of Mechanical and Industrial Engineering at the Norwegian University of Science and Technology. He does research in early-stage product development, focusing primarily on prototyping, the role of makerspaces and open-source hardware. His research is part of TrollLabs, a Prototyping and product development lab/community at NTNU in Trondheim, Norway.

Marius Auflem currently works at Laerdal Medical as an industrial PhD candidate through the Department of Mechanical and Industrial Engineering at the Norwegian University of Science and Technology. Marius does research in early-stage Product Development in the context of medical training and simulation equipment. His research is part of TrollLabs, a Prototyping and product development lab/community at NTNU in Trondheim, Norway.

Oscar Lilleløkken is a mechanical engineering master student at the Department of Mechanical and Industrial Engineering at the Norwegian University of Science and Technology. He is doing his master thesis on prototyping medical training equipment concepts. The research is part of TrollLabs, a Prototyping and product development lab/community at NTNU in Trondheim, Norway.

Martin Steinert is a professor of engineering design and innovation in the Department of Mechanical and Industrial Engineering at the Norwegian University of Science and Technology (NTNU). He teaches fuzzy front-end engineering for radical new product/service/system concepts and graduate research seminars for PhD students engaged in topics related to new product design and development. His various research projects are usually multidisciplinary (ME/CS/EE/Neuro- and cognitive science) and are often connected with the industry. The aim is to uncover, understand, and leverage early-stage engineering design paradigms with a special focus on human-machine/object interactions.

References
ABSIM. 2020. “ACDET-ABSIM | Abdominal Simulator | USA.” Accessed 1 June 2020. https://www.acdet-absim.com/.
Arita, R., T. Numata, N. Saito, S. Takayama, T. Togashi, S. Kaneko, A. Kikuchi, et al. 2019. “Development of a Medical Education Program with Abdominal Palpation Simulators to
Support the Understanding of Traditional Japanese (Kampo) Medicine in Beginners.” *Traditional & Kampo Medicine* 6 (3): 148–155. doi:10.1002/tkm2.1230.

Aubin, A., K. Gagnon, and C. Morin. 2014. “The Seven-Step Palpation Method: A Proposal to Improve Palpation Skills.” *International Journal of Osteopathic Medicine* 17 (1): 66–72. doi:10.1016/j.jiosm.2013.02.001.

Auflem, M., J. F. Erichsen, and M. Steinert. 2019. “Exemplifying Prototype-Driven Development through Concepts for Medical Training Simulators.” *Procedia CIRP* 84: 572–578. doi:10.1016/j.procir.2019.04.202.

Blandford, A., D. Furniss, and C. Vincent. 2014. “Patient Safety and Interactive Medical Devices: Realigning Work as Imagined and Work as Done.” *Clinical Risk* 20 (5): 107–110. doi:10.1177/1356262214556550.

Cooper, J. B., and V. R. Taqueti. 2004. “A Brief History of the Development of Mannequin Simulators for Clinical Education and Training.” *BMJ Quality & Safety* 13 (1): 11–18. doi:10.1136/qshc.2004.009886.

Dede, C. 2009. “Immersive Interfaces for Engagement and Learning.” *Science* 323 (5910): 66–69. doi:10.1126/science.1167311.

Dieckmann, P., M. Patterson, S. Lahlou, J. Mesman, P. Nyström, and R. Krage. 2017. “Variation and Adaptation: Learning from Success in Patient Safety-Oriented Simulation Training.” *Advances in Simulation* 2 (1): 21. doi:10.1186/s41077-017-0054-1.

Dinsmore, M., N. Langrana, G. Burdea, and J. Ladeji. 1997. “Virtual Reality Training Simulation for Palpation of Subsurface Tumors.” Paper presented at the IEEE 1997 Annual International Symposium on Virtual Reality, Albuquerque, USA, March 1–5.

Ege, D., O. Lilleløkken, M. Auflem, and M. Steinert. 2020. “Dealing with Ecological Validity and User Needs When Developing Simulation Based Training Equipment – Case Study of a Medical Palpation Task Trainer.” *Procedia CIRP* 91: 722–727. doi:10.1016/j.procir.2020.03.116.

Escobar-Castillejos, D., J. Noguez, L. Neri, A. Magana, and B. Benes. 2016. "A Review of Simulators with Haptic Devices for Medical Training." *Journal of Medical Systems* 40 (4): 104. doi:10.1007/s10916-016-0459-8.

Hamstra, S. J., R. Brydges, R. Hatala, B. Zendejas, and D. A. Cook. 2014. “Reconsidering Fidelity in Simulation-Based Training." *Academic Medicine* 89 (3): 387–392. doi:10.1097/ACM.0000000000000130.

Houde, S., and C. Hill. 1997. “What Do Prototypes Prototype?.” In *Handbook of Human-Computer Interaction* (Second Edition), edited by Marting G. Helander, Thomas K. Landauer, and Prasad V. Prabhu, 367–381. Amsterdam: North-Holland. doi:10.1016/B978-044481862-1.50082-0.
Issenberg, S. B., W. C. Mccaghie, E. R. Petrusa, D. Lee Gordon, and R. J. Scalese. 2005. “Features and Uses of High-Fidelity Medical Simulations That Lead to Effective Learning: A BEME Systematic Review.” Medical Teacher 27 (1): 10–28. doi:10.1080/01421590500046924.

Laerdal. 2021. “Abdominal Examination Trainer.” Accessed 11 January 2021. https://laerdal.com/no/products/skills-proficiency/limbs-things/abdominal-examination-trainer/

Lawton, R., R. R. C. McEachan, S. J. Giles, R. Sirriyeh, I. S. Watt, and J. Wright. 2012. “Development of an Evidence-Based Framework of Factors Contributing to Patient Safety Incidents in Hospital Settings: A Systematic Review.” BMJ Quality & Safety 21 (5): 369–380. doi:10.1136/bmjqs-2011-000443.

Leape, L. L., T. A. Brennan, N. Laird, A. G. Lawthers, A. R. Localio, B. A. Barnes, L. Hebert, J. P. Newhouse, P. C. Weiler, and H. Hiatt. 1991. “The Nature of Adverse Events in Hospitalized patients. Results of the Harvard Medical Practice Study II.” The New England Journal of Medicine 324 (6): 377–384. doi:10.1056/NEJM199102073240605.

Lilleløkken, O., D. Ege, M. Auflern, and M. Steinert. 2020. “Proposing a Novel Approach for Testing Complex Medical Task Trainer Prototypes.” Paper presented at NordDesign 2020, Lyngby, Denmark, August 12–14.

Money, A. G., J. Barnett, J. Kuljis, M. P. Craven, J. L. Martin, and T. Young. 2011. “The Role of the User within the Medical Device Design and Development Process: Medical Device Manufacturers’ Perspectives.” BMC Medical Informatics and Decision Making 11 (1): 15. doi:10.1186/1472-6947-11-15.

Motavalli, A., and D. Nestel. 2016. “Complexity in Simulation-Based Education: Exploring the Role of Hindsight Bias.” Advances in Simulation 1: 3. doi:10.1186/s41077-015-0005-7.

Norman, G., K. Dore, and L. Grieerson. 2012. “The Minimal Relationship between Simulation Fidelity and Transfer of Learning.” Medical Education 46 (7): 636–647. doi:10.1111/j.1365-2923.2012.04243.x.

Reay, S. D., G. Collier, R. Douglas, N. Hayes, I. Nakarada-Kordic, A. Nair, and J. Kennedy-Good. 2017. “Prototyping Collaborative Relationships between Design and Healthcare Experts: Mapping the Patient Journey.” Design for Health 1 (1): 65–79. doi:10.1080/24735132.2017.1294845.

Reid-McDermott, B., M. Browne, D. Byrne, P. O’Connor, E. O’Dowd, C. Walsh, C. Madden, and S. Lydon. 2019. “Using Simulation to Explore the Impact of Device Design on the Learning and Performance of Peripheral Intravenous Cannulation.” Advances in Simulation 4 (1): 27. doi:10.1186/s41077-019-0118-5.

Ribeiro, M. L., H. M. Lederman, S. Elias, and F. L. S. Nunes. 2016. “Techniques and Devices Used in Palpation Simulation with Haptic Feedback.” ACM Computing Surveys 49 (3): 1–48. doi:10.1145/2962723.

Rieber, L. P., and D. Noah. 2008. “Games, Simulations, and Visual Metaphors in Education: Antagonism between Enjoyment and Learning.” Educational Media International 45 (2): 77–92. doi:10.1080/09523980802107096.

Rowe, A., M. Knox, and G. Harvey. 2020. “Re-Thinking Health through Design: Collaborations in Research, Education and Practice.” Design for Health 4 (3): 327–344. doi:10.1080/24735132.2020.1841918.

Scerbo, M. W., and S. Dawson. 2007. “High Fidelity, High Performance?” Simulation in Healthcare 2 (4): 224–230. doi:10.1016/S1170-2101(07)60003-2.

Steinert, M., and L. J. Leifer. 2012. “Finding One’s Way: Re-Discovering a Hunter-Gatherer Model Based on Wayfaring.” International Journal of Engineering Education 28 (2): 251.
Stokes-Parish, J. B., R. Duvivier, and B. Jolly. 2020. “How Does Moulage Contribute to Medical Students’ Perceived Engagement in Simulation? A Mixed-Methods Pilot Study.” Advances in Simulation 5 (1): 23. doi:10.1186/s41077-020-00142-0.

Takeuchi, H., and I. Nonaka. 1986. “Stop Running the Relay Race and Take up Rugby.” Harvard Business Review 64 (1): 137–146.

Thomke, S., and D. Reinertsen. 1998. “Agile Product Development: Managing Development Flexibility in Uncertain Environments.” California Management Review 41 (1): 8–30. doi:10.2307/41165973.

Ullrich, S., and T. Kuhlen. 2012. “Haptic Palpation for Medical Simulation in Virtual Environments.” IEEE Trans Vis Comput Graph 18 (4): 617–625. doi:10.1109/TVCG.2012.46.

Wilson, E., D. G. Hewett, B. C. Jolly, S. Janssens, and M. M. Beckman. 2018. “Is That Realistic? The Development of a Realism Assessment Questionnaire and Its Application in Appraising Three Simulators for a Gynaecology Procedure.” Advances in Simulation 3 (1): 21. doi:10.1186/s41077-018-0080-7.

Yakubo, S., Y. Kinoshita, T. Aki, and H. Ota. 2008. “Improvement of a Simulator Production Project for Abdominal Palpation in Kampo Medical Training.” Kampo Medicine 59 (4): 595–600. doi:10.3937/kampomed.59.595.

Zenios, S., J. Makower, P. Yock, T. J. Brinton, U. N. Kumar, L. Denend, and T. M. Krummel. 2009. Biodesign: The Process of Innovating Medical Technologies. 1st ed. Cambridge, UK; New York, NY: Cambridge University Press.

Zhang, J., T. R. Johnson, V. L. Patel, D. L. Paige, and T. Kubose. 2003. “Using Usability Heuristics to Evaluate Patient Safety of Medical Devices.” Journal of Biomedical Informatics 36 (1–2): 23–30. doi:10.1016/S1532-0464(03)00060-1.