Pre-publication versions of the papers

British HCI Conference 2021

Post-Pandemic HCI – Living digitally

Day 2: Wednesday 21 July 2021

Warning and apology: Please note that the formatting of papers is not perfect in this document and does not reflect how they will appear in the final conference proceedings!

Table of Contents

Session 5. HCI for healthcare. Session chair: Anirudha Joshi

Future outlook on the materialisation, expectations and implementation of Digital Twins in healthcare

Christel De Maeyer
Department of Industrial Design
Eindhoven University of Technology
c.a.a.d.maeyer@tue.nl

Panos Markopoulos
Department of Industrial Design
Eindhoven University of Technology
p.markopoulos@tue.nl

In this paper we examine how experts in digital transformation of healthcare envision the application of Digital Twins. The concept of a Digital Twin refers to a digital replica of potential and actual physical assets, processes, people, places, systems and devices that can be used for various purposes including scientific experiments, simulations and prediction of intervention outcomes. Digital Twins are an emerging technological vision and while evocative as a term it holds different promises and connotations for different application areas and may evolve in very different directions. In order to examine how these directions can develop and impact healthcare we used the Delphi method to reach a consensus among experts on three different research questions we have put forward, namely how experts see the materialisation, expectations and implementation of a Digital Twin in healthcare. Our main conclusion is that Digital Twins are seen as enabling preventive healthcare and trial-and-error approaches to support personalised medicine and/or patient centred care.

Digital Twins, personalised healthcare, personal informatics, preventive healthcare

1. INTRODUCTION

The concept of a Digital Twin is gaining traction in various fields including medicine, logistics and engineering, in reference to a digital replica of potential and actual physical assets, processes, people, places, systems and devices. These replicas can be used for various purposes, including the analysis of existing situations but also the prediction of intervention outcomes. With the abundance of data available over people's life and daily living the prospect emerges of digital twins in healthcare, which can potentially help understand and diagnose different pathologies, simulate therapies or medical treatments, and predict their outcomes. Such predictions could be made for an individual or a whole population. This prospect is a natural progression from trends in self-tracking technologies and especially the recent increase in digitization of everyday life that has followed the pandemic of covid-19, pertaining to applications for monitoring contagion in forms of contact tracing apps (Rodríguez, 2021), (Lewis, 2021), controlling confinement and social distancing by monitoring the density of cities and regions, using Smart City tools. (World Economic Forum, 2020).

The realization of Digital Twins in healthcare is closely related to the increasing availability of personal health data, as people can easily gather data about themselves, keeping track of different biomarkers, physical activities, and wellbeing in general. Interconnected sensing devices and appliances at home and the workplace enable tracking of an individual's environment and health related behaviours and parameters. Integrated electronic health records provide caregivers and patients easy access to records on vaccinations, medication, and population studies. Scholars are speaking of humans akin to "walking sensor platforms" (Smith, 2016) or "Data selves" (Lupton, 2019). Technically Digital Twins in healthcare require the integration and interoperability of different sources of information, e.g., personal tracking devices and electronic health records. While technically feasible, full integration that could support digital twinning remains a challenge. For example, scholars researching the Quantified Self have been pointing out that the data one gathers with personal devices are not interconnected with each other, there is not really interoperability between all the devices and their accompanied software and by consequence, there is no context surrounding the data (Fuller, 2020), (Elayan, 2021) (Schwartz, 2020).

Next to technical integration challenges, there are major concerns regarding the protection and security of the data that a Digital Twin might hold, what this data could entail and how its availability may affect healthcare systems and societies at large. Continuous monitoring can lead to data-surveillance, which points to a need for changing health policies and governance: "There should be a major effort to democratizing technology", working on more inclusive policies and regulations (Rasheed, 2020, p. 18), Fukuyama, talks about 'human enhancements' and the potential to create a new class system, "the fear that human enhancement technologies might lead to different classes of people, and therefore have a disruptive effect on our democratic institutions." (Fukuyama, 2002). Various authors argue how individuals should have autonomy and agency to work with their Digital Twin, "These systems still raise the questions of autonomy and the potential development of nanny technologies which prevent humans from working through their own problems unfettered". (Bartneck, 2020, p. 75). Today, a Digital Twin could be materialized by leveraging data that is already in abundance. It becomes important to assess what the systemic challenges are, pertaining to the realization of digital twins in healthcare: How are we going to use the data? What purposes or expectations are there, and how do we make all the correlations work (Malakuti, 2018)? Can we do this in a safe environment that complies to the guidelines for humane Artificial Intelligence (AI HLEG, 2019)? What are the expectations of a Digital Twin in healthcare and how should we implement a Digital Twin in healthcare?

As Digital Twins are still a future vision, these questions are rather speculative and how they will be resolved depends on the pursuits and sustained efforts of related stakeholders and interest groups. In an attempt to understand the direction in which Digital Twins can evolve, we set out to investigate how stakeholders approach these issues, focusing on three aspects of a Digital Twin in healthcare: their materialisation, the expectations from them and the implementation of a Digital Twin in healthcare especially considering ethics. We report on a Delphi study with a panel of experts (N=9) from the healthcare sector in Belgium, which was conducted between July-October 2020. Our study included three rounds of questions as above on the materialisation, expectations, and implementation of a Digital Twin in healthcare systems.

The Delphi study that we report in the following sections of this concluded that a) Implementing Digital Twins in healthcare can bring value mostly in relation to short clinical trials on Digital Twins b) Experts

expect that Digital Twins will give way to personalised medicine for patients c) Implementation in healthcare requires including a human in the loop for all decision making and usage of the Digital Twin.

In the remainder of the paper, we briefly overview the state of the art regarding Digital Twins in healthcare and we elaborate on the aims of this research the Delphi methodology we followed and our findings.

2. WHAT ARE DIGITAL TWINS?

A Digital Twin refers to a digital replica of potential and actual physical assets (physical twin), processes, people, places, systems, and devices that can be used for various purposes (Grieves M. , 2014). A virtual representation of a physical object holds a lot of detail and is a rich representation of the physical object, almost indistinguishable from the original (Grieves M. , 2014). The concept of a Digital Twin was introduced in 2003, at the University of Michigan during a course on Product Lifecycle Management as a tool or a system for tracking and documenting the process from the production of raw materials all the way to finished products. The automation of production, enables collecting a wealth of data from production lines, giving insights on the design and maintenance of a physical product, using disparate data sources, such as physical non-destructive sensing technologies, including sensors and gauges, lasers and so forth. The notion of a Digital Twin concept that could illustrate a physical object and its detailed underlying information to come to a closed production loop, could thus give an overview on its production cycle, enable optimization and a more cost-effective production line (Grieves M. , 2014).

A Digital Twin must fulfil three prerequisites: first one needs a physical object or asset that we want to twin, second, one needs a virtual version of the physical object and third one needs to establish and maintain the connections between the physical and virtual object through data collection bringing everything together in the virtual representation. The virtual side of a Digital Twin allows testing and/or simulating, emulating, and mirroring different situations before these are applied upon the physical object or entity. This enables the assessment of consequences before an actual situation occurs. In health-related applications this offers the possibility to forecast and/or predict the outcome of certain medical treatments and even to personalize them towards the patient. As mentioned in the white paper by Grieves (Grieves M. , 2014), one could have 'lightweight versions' of a virtual model. This allows physicians to look at different characteristics of the patient and isolate different pathologies to monitor and control for a certain patient, considering the different aspects and characteristics of a human being as seen in figure 1. At a systemic level Digital Twins can potentially facilitate a transition towards a more care oriented medical approach instead of the cure approach we have today. 'A Digital Twin is meant to be a living, intelligent and evolving model depending on data and information it gets from the physical counterpart' (Barricelli, 2019, p. 167656). The actual Digital Twinning process is possible because of the continuous synchronization, communication and interaction between the physical entity and its virtual replica, a 'closed-loop optimization' as it were. Foremost because the affordance of exchanging real-time data between the physical object and its virtual counterpart and the capability of big data storage nowadays (Barricelli, 2019). However, to get more clarity on the different manifestations of a Digital Twin, Grieves (Grieves M. V., 2017) elaborates the definitions of the different manifestations of a Digital Twin. The first definition we discussed above is the general definition of a Digital Twin (DT or DTP). The different manifestations of a Digital Twin exist of two types: a Digital Twin Instance (DTI) and a Digital Twin Environment (DTE) as shown in figure1.

As shown in the visual above, a 'Digital Twin prototype (DTP)' holds the necessary information to model a twin or virtual version of the physical object. The Digital Twin prototype holds information of the physical object. In the case of healthcare, data gathered by patients and different stakeholders (EPR, lifestyle and environmental data), the data is annotated and comes with a set of requirements together with a bill of materials to use a Digital Twin and how to apply different possible scenarios as we will see further on in this paper, e.g., short clinical trials, which will be launched on a Digital Twin instead of on a patient. In addition, there is a bill of disposal connected to the Digital Twin prototype, in the case of healthcare these bills are regulated by the government in most cases, e.g., when a patient dies the Digital Twin might be archived for x years, or patients can give consent to use it for future research.

A Digital Twin Instance (DTI) allows for interrogation and trial and error testing on the DTP. For example, depending on certain scenarios and use cases, surgeons could simulate a knee operation and forecast the revalidation time a patient will need to recover looking at the medical and lifestyle data (historical and current data) of that patient. In addition, there is the Digital Twin Environment (DTE). This is a multi-domain physics application space for operating on Digital Twins that allows for prediction of future behaviour and performance of the physical object, in our case the human (Grieves M. V., 2017, p. 94).

2.1 Digital Twin in healthcare

The notion of a Digital Twin in healthcare refers to idea that we have a digital replica of a human being, that will adapt in real-time to the data that nurtures it. As people accumulate data about themselves in different settings through sensory devices in their home environment, biomarkers delivered by wearable devices and medical data that is digitized in hospitals, deep learning techniques can be used to analyse lifestyle patterns, one can monitor the positive and negative consequences of a certain lifestyle, spot risks or anomalies before they arise. Contextual data such as emotional states, environmental data, age and preferences can potentially enable holistic considerations of individuals (Fuller, 2020), (Saddik, 2018). Even though they are still at an early stage, we see already commodity applications that monitor biomarkers of individuals. Applications such as Fibricheck (Fibricheck, sd), for remote heart rhythm monitoring are medically approved and have been used in clinical trials, for prevention purposes or as (remote) following up after therapy.

Another notable application that can be thought of as an early example of a Digital Twin, is Epicare@home (Byteflies, sd), which is a personalised wearable allowing seizure monitoring. Epicare@home bridges between the home of the patient and the hospital. Clinical data is collected by the patient in the comfort of their home through the wearable and shared with the neurologist. Epicare@home enables personalised treatments for epilepsy patients.

Remote surgery is another potential application area, where surgeons could remotely plan and simulate a surgery, minimizing risks on life. 'In the future the patient and the surgeon might also possess their own Digital Twin which would enable the system to read historical data and real time health and performance data. This leads us to a situation where several Digital Twins are brought together into one in order to accomplish a task. The Digital Twins needs to be fired up to bring all the required agents together. This means that Digital Twins will be born and terminated in relation to specific events, which further magnifies the security issues mentioned before' (Lakki, 2019, p. 20334), (Barricelli, 2019).

3. A DELPHI STUDY ON THE FUTURE OF DIGITAL TWINS IN HEALTHCARE

The realization of a Digital Twin is a process that requires the synergy of diverse stakeholders in different sectors, who need to align their efforts to overcome diverse technological, business, and societal challenges. To gauge the challenges and opportunities ahead and how best to approach the vision of the Digital Twins in healthcare, we set out to solicit and synthesize the views from experts active in this area. We set up a panel of Belgian experts in digital technology for healthcare, to examine the materialisation, expectations, and implementation of Digital Twins in healthcare, considering the ethical guidelines that are in place for AI systems, as these have been defined by EU experts (AI HLEG, 2019). Anticipating that experts approach the concept of Digital Twins from different angles, we sought a method to help solicit a consensus view from them. We used the Delphi research method (Helmer, 1950-60) to structure our inquiry and come to a consensus on aforementioned topics.

3.1 The Delphi method

The Delphi method is a structured approach for soliciting and combining input from experts in pursuit of a consensus view (Renzi, 2015; Yang, 2019). In this method the panel is assembled purposefully, to include a diversity of experts who can provide sufficient variety of insights and viewpoints. Further, The Delphi method allows for three types of questions where different expertise and experts are needed (Gordon, 1994):

- Forecast the occurrence of future developments. For example, 'the future of virtual agents in healthcare'
- Desirability of a certain development. For example, 'designing designer babies with CRIPR technology'
- The means for achieving or avoiding a future state. Involves policies concerning implementations. The who, what and when and where, needs to be tight to the objectives sought and the likelihood that a policy will accomplish its goals. For example, 'how to define policies for mass surveillance mechanisms'

The Delphi methodology may combine different methods for questioning the expert panel, both qualitative and quantitative methods, such as in-depth interviews, group interviews or online surveys. Experts remain anonymous to each other to allow them to evaluate each other's opinions free from

social desirability bias. Researchers are tasked with surveying opinions, collecting, and organizing them to facilitate voting, aiming to converge at a consensus view. The method involves surveying opinions, pooling them together and iterative steps of ranking, and rating these opinions to reach a consensus among the participants. To avoid biases relating to group influences, there are no discussions among the participants during the iterative process of this method.

Over the years digital media have been increasingly applied in Delphi studies to share contributions efficiently, to ensure anonymity of input or even the asynchronous execution of Delphi studies using various groupware.

3.2 Participants

We applied purposive sampling, to form a panel of experts with knowledge and expertise regarding future healthcare, in different domains that are interconnected with each other. We received a list of experts that collaborate with the government, and more specifically with the department of Economy, Innovation and Science, as this department is also responsible for innovation funding and setting up innovation hubs in Belgium, fostering triple helix collaborations (Etzkowitz, 2017). From that list we chose our experts (N9) and recruited from researcher's own network. A typical Delphi panel has about 8 to 12 members, other sizes are possible but a minimal of 7 is advised to avoid deteriorate accuracy. (Mullen, 2003)

Each candidate received a recruiting mail, and an intake conversation was done with each expert. There were no dropouts during this research and panel experts didn't have contact with each other.

3.3 Materials

The Qualtrics software was used through the entire project, except for Q1 was done in Google forms.

To kick off the study, and to create a common frame of reference for the experts, we started with viewing a publicly available and widely disseminated explanation video [1] of a TEDX talk, a local independent conference based on TED conference format, featuring a talk by Jacqueline Alderson, Associate Professor, at Auckland University of Technology introducing the concept of Digital Twins. The video was chosen because it captures an academic perspective and is more neutral when compared to related corporate vision videos that represent a specific business interest. The video served as a preamble for the first question to the panel regarding the materialisation of a Digital Twins in healthcare.

Below we elaborate on the setup of the questionnaire and the schedule or timing of the different questions on the materialisation, expectation, and implementation of a Digital Twin in healthcare.

***Table 1**: Overview of experts*

Disciplines or skills	Organizations
Health Tech and Bio Tech engineering. Guidance for health tech start-ups. Academic research in the domain of healthcare, IT managerial tasks in University Hospital	Innovation Hub with foci on digital healthcare transformation *(2 experts) (Academic and Industrial appointment)*
Bioengineer, passion for health and health care and a sweet spot for science	*Health House, with focus on the future of healthcare based on scientifically validated content and the impact technology will have on this future. (Academic and Industrial appointment)*
Exec in Biotech companies, academic cancer, omics and digital health adoption	*Healthcare futurist Academic and (Industrial appointment)*

[1] https://www.youtube.com/watch?v=oQh8ugLWrpU

research, CEO Digital health agency, cochair Digital Twin Consortium	
CEO of a private retirement home. Today guidance in health tech start-ups	*Retirement home and start-up guidance in health tech (Industry)*
Strategic and policy vision of the continuum of life, cure, care, comfort & wellbeing	*Caregiving implementor, platform for caregivers (Industry)*
Creative & strategical marketing & communication agency for the silver market	*Marketing Bureau for seniors (Industry)*
Product management with demonstrated history of working in healthcare	*Medical Imaging and workflows (Industry)*
Specialized in digital transformation (international and national)	*Caregiving implementor for caregivers and patients (Industry)*

3.4 Procedure

After selecting the experts, we launched a cycle of the three iterations we were interested in, each addressing one of the main research questions. The research questions were developed based on the existing literature in the area of Digital Twins in healthcare with a focus on 'Forecast on the occurrence of future developments' (Gordon, 1994).

1. How do you see the materialisation of a Digital Twin in healthcare?
    a. Experts gave their ideas and thoughts on the materialisation of a Digital Twin in healthcare.
    b. The input was synthesized and ranked by importance.
    c. Kendall's W was used to measure the agreement and consensus of Q1.
    d. We selected the top three ranked items and asked the participants to rate their agreement to the items on a 7-point Likert scale (1- strongly disagree to 7- strongly agree). This step was implemented to get a focused and prioritized view of experts on all the concepts that were mentioned in relation to the materialisation of a Digital Twin.
2. What are the expectations of a Digital Twin in healthcare?
    a. Experts received a list of expectations based on EU guidelines, 'from fundamental rights to ethical guidelines' (AI HLEG, 2019, p. 10).
    b. From the input list, experts were asked to rate the input list according to desirability on a Likert scale (1 undesirable – 5 extremely desirable)
    c. This list was then ranked to come to Top 3.
3. How do you see the implementation of a Digital Twin in healthcare?
    a. Experts gave their ideas and thoughts on the implementation of a Digital Twin in healthcare.
    b. The synthesized input list was then ranked by importance by the experts.
    c. Kendall's W was used to measure the agreement and consensus of Q3.
    d. After ranking the synthesized input list, experts were asked to rate the Top 3.

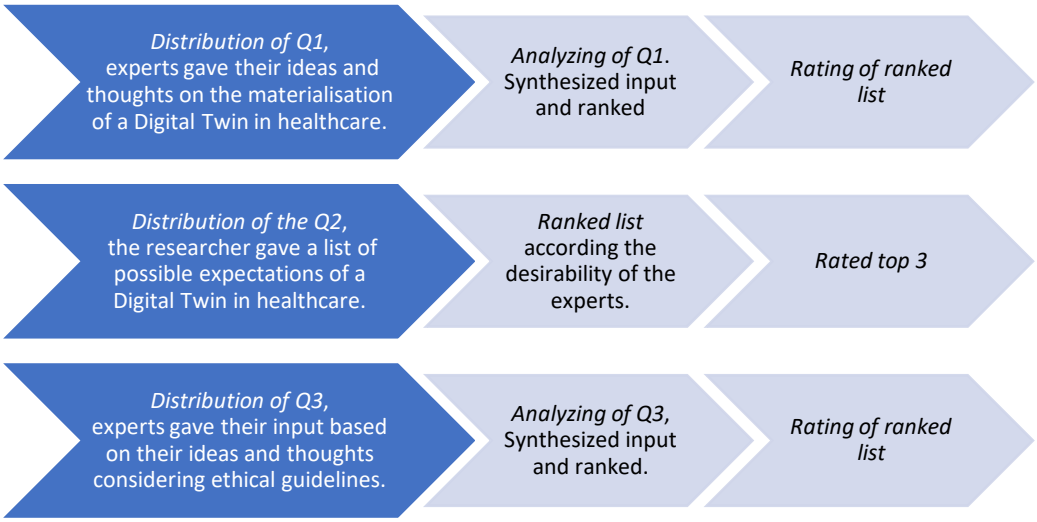


***Figure 1**: Setup of the distribution of the questionnaire using the Delphi method*

3.4 Timeline of study

***Table 2**: Timeline of distribution research questions*

June	Recruiting of experts
7.7.2020	*First research question – Listing by experts, the materialisation of a Digital Twin in healthcare.*
22.8.2020	*Ranking of synthesized input first question*
10.9.2020	*Rating of the ranking first question (top 3).*
10.9.2020	*Ranking of second question (input given by the researcher), the expectations of a Digital Twin in healthcare.*
21.9.2020	*Rating of input on second question (top 3)*
27.9.2020	*Third research question – listing by expert, the implementation, considering ethical guidelines of a Digital Twin in healthcare.*
11.10.2020	*Ranking of the input on third question.*
19.10.2020	*Rating on the third question (top 3)*
20.10.2020	*Closing*

4. RESULTS

4.1 Q1. How do you see the materialisation of a Digital Twin?

Experts were asked to fill in their thoughts and ideas on the materialisation of a Digital Twin in healthcare. Overall, the answers to this question went in the direction of preventive care and shifting the focus of healthcare on care rather than cure. Experts suggested the notion that a Digital Twin could be an environment used for trial and error of different interventions, based on historical and current data of a patient. Furthermore, on a macro level Digital Twins could be used to reorganize our healthcare system as we know it. In terms of health insurance,

it could create a health index of a certain population. At a micro level, Digital Twins could lead towards more precise and personalised medicine.

By grouping and collapsing similar ideas and rephrasing them in a consistent way, we composed a list of 11 concepts regarding the possible promise of Digital Twin. In round two we asked experts to rank

the items of this list (1 is most important, 11 is least important). The items and their average ranking are shown in table 3.

***Table 3**: Topics Ranked Order of materialisation of Digital Twin*

	M (STD)
Digital Twin as a concept for trial and error in healthcare instead of operating trial and error directly on patients.	*2.56 (2.27)*
Digital Twins as facilitators towards preventive healthcare	*2.56 (0.96)*
Digital Twin in Intensive Care Units where patients are followed in detail, based on certain parameters (controlled lab environment) might be a good starting point	*4.44 (2.91)*
Digital Twin with a focus on (a) certain parameter(s) within a controlled environment for now.	*5.33 (1.83)*
Digital Twin as 24/7 personal assistant where the whole chain of family, healthcare givers and trusted network can follow the journey of a patient or ageing person.	*5.78 (2.62)*
Digital Twin will help with effective care interventions	*4.78 (1.81)*
Digital Twins as facilitator towards longevity	*7.22 (2.35)*
Digital Twin could have potential to save money on our healthcare in general but also towards the patient (patients who are really taking care of themselves should be rewarded), those who are not complying will have to pay more for insurance or consultations.	*7.44 (2.67)*
Digital Twin as a prevention healthcare based on a health index. Linked to our social healthcare system (insurance)	*7.67 (2.45)*
Digital Twins could avoid loneliness, solitude	*8.67 (2.45)*
Digital Twin as prevention maintenance is still far away	*9.56 (1.77)*

Kendall's W coefficient of concordance was applied to verify the agreement level of the ranking process. The level of agreement of the list was low. Kendall's W=0.493, though significant (p = 0,000). Examination of the data showed that few ideas were aligned with the experts.

We selected the top three ranked responses and asked them to rate their importance, leading to a slight reordering, as presented in Table 4, compared to the original ranking of the full collection of items.

***Table 4**: Importance ratings for the three top ranked items from Q1 regarding the materialisation of a Digital Twin Likert-scale (1- strongly disagree to 5- strongly agree)*

	M (STD)
Digital Twin as facilitator towards preventive healthcare	*4.33 (0.47)*
Digital Twin in intensive care units where patients are followed in detail, based on certain parameters (controlled lab environment) might be a good starting point)	*3.89(0.74)*
Digital Twin as a concept for trial and error in healthcare instead of operating trial and error directly on patients	*3.67(1.15)*

Regarding prevention, a Digital Twin could create more awareness with individuals based on the bio- and lifestyle feedback, individuals could adjust their behaviour if needed. Furthermore, Digital Twins could give the ability to forecast certain scenarios in life. Such scenarios have been discussed before in the context of quantified self-applications, for example, a marathon runner, who prepares for

marathons together with his sports doctor; the doctor advises which diets to follow, which exercises are good to prepare to avoid injuries based on past data and current lifestyle of the marathon runner (De Maeyer, 2020), (Bagaria, 2020). Another application scholars foresee, is simulating side effects of certain drugs/medication (Fuller, 2020), (Bruynseels, 2018).

As one of the experts stated 'Prevention: it would be fantastic if, in addition to a health index, your Digital Twin would also provide certain risk scores for certain chronic diseases and the course of those scores over time'.

The second most promising idea for experts, concerns the notion of Digital Twins used in intensive care as a controlled environment to follow patients very closely based on certain/specific parameters. As one of the experts wrote: 'the state of the art "Digital Twin" environment is an ICU in which patients (in a sick context) are optimally and continuously monitored based on numerous parameters. However, this is a perfectly controlled environment, it seems interesting to us to start from there in order to gain ideas that you can translate to a healthy population (including healthy elderly people)'.

The third rated idea concerned the notion of trial-and-error, a mechanism that can be applied on the Digital Twin replica of the patient before applying clinical trials to a certain patient. These trial-and-error mechanisms that could be applied to a Digital Twin could evolve towards personalised medicine or human (patient)-centred healthcare (Rasheed, 2020), (Searl, 2010). Clinical trials could be done on 'Lightweight versions of the virtual model which exist on a temporary basis for these tests for example (Grieves M. , 2014), as we see in the applications of Fibricheck or Epicare@home.

4.2 Q2 What are the expectations of a Digital Twin, input given by researcher

The expectations from Digital Twins in healthcare as discussed in the EU guidelines, (AI HLEG, 2019) were rated for their desirability in two steps. First experts each provided their rating of desirability. The list is shown in table 5 ordered from the most to the least desirable. Then we asked experts again to rate the top three items; the ratings are shown in table 6.

***Table 5**: Ratings of desirability on Q2 What are the expectations of a Digital Twin? Likert-scale (1- undesirable to 5- extremely desirable)*

	M (STD)
Getting insights on remedial process for a specific patient (personalised medicine)	*4.44(0.50)*
Getting context around the data to do a more accurate assessment and support decision making.	*4.33(0.47)*
Getting oversight on the health status of a population	*3.63(0.48)*
Getting insights in a longitudinal data-driven lifecycle of individuals	*3.56(0.83)*
Getting insights in vulnerable groups, pro-active remediation towards a more equal society?	*3.25(0.43)*
Digital Twins in healthcare will be an on-demand service for individuals?	*3.13(0.60)*
Health insurance will be reinvented, they will work with Health KPI's which are parameters for your insurance	*3.00(1.00)*

***Table 6**: Ranked top 3 on the question 'What are the expectations of a Digital Twin'? According to desirability*

	M (STD)
Getting insights on remedial process for a specific patient (personalised medicine)	*1.86 (0.64)*
Getting context around the data to do a more accurate assessment and support decision making.	*2.00 (0.93)*
Getting oversight on the health status of a population	*2.14 (0.83)*

Regarding the top 3 expectations of a Digital Twin in healthcare, we see that 'Getting insights on remedial process for a specific patient (personalised medicine)' is connected to the last aspect of materialisation of a Digital Twin. The trial and error on a Digital Twin could lead way and give insight as well on a 'remedial process' for a specific patient, hence leading towards 'personalised medicine'.

The second expectation 'Getting context around the data to do a more accurate assessment and support decision making'. One of the expert's input: 'These data are for individual use (each user is unique), collected anonymously so that new connections can be discovered via machine learning, and the care is further optimized compared to what is known today (both medicinal and non-medicinal). In the long term I see that dashboard as a little less static but effective as a virtual replica of the user where data 'flows' through that virtual body in real-time or near-real time. The timeframe can then be moved or stretched via a slider, so that the evolution over time of a specific parameter becomes visible'. The interoperability of the data silos that are there today, will become extremely important to create context around data- driven healthcare if one wants to use the context to support decision making and get more accurate data-driven information (Barricelli, 2019), (Rivera, 2019). Indeed, the last rated expectation could give a health status on a certain or given population. The pitfalls here are the biases that are generated within a data-driven model. There needs to be attention to these biases as they can lead to new health policies, exclude or include a certain privileged or underprivileged population (O' Neill, 2016). This connects directly to the next research question concerning the implementation of a Digital Twin, considering ethical guidelines set out by an expert group on EU level.

4.3 Q3 How do you see the implementation of ethically guidelines?

Experts gave their opinions and thoughts regarding the implementation of ethical guidelines. These thoughts were then shared anonymously between experts who rank ordered them by importance. The individual rankings were merged into the synthesized rank ordering of table 7.

***Table 7**: Synthesized ranked input from experts on the question 'How do you see the implementation of ethically guidelines?*

	M (STD)
Digital Twins are in fact augmenting human capabilities. That implies that humans decide at the end guided by algorithms, medical professionals, and doctors GP's.	*2.22 (1.99)*
A medical professional should always be in the loop and a Digital Twin-based solution is only a tool for a medical professional to aid in the process of decision making	*3.33 (1.76)*
Complete transparency on the how and why we use a Digital Twin should give trust to the patient to use a Digital Twin	*3.44 (1.57)*
There will always be a human in the loop - a medical professional - to ensure that the advice/treatment inspired by the Digital Twin is safe for the patient.	*3.78(1.31)*
Giving the possibility to patient to decide which data could be added to the Digital Twin. Patient is to decide whether or not to use Digital Twin for a specific application. Patient could stop usage of Digital Twin at any time.	*4.67 (1.70)*
A patient should always accept that whatever data has been put in by him/her this data will be anonymized and used for the good of the crowd. (Like a doctor today, uses all her/his insights of her/his patients to finetune to her/his other patients)	*5.11 (1.28*
Focus on educational aspects - explain citizens what 'personalised medicine' even means - Based on that we can engage people to start sharing personal data useful to build a DT	*5.44 (2.11)*

We applied a Kendall W coefficient of concordance to measure the agreement and consensus, in this case the agreement was very low, W= 0,276, p-value = 0,021.

4.6 Implementation list Top 3 ranked

***Table 8**: Ranked top 3 on the question 'How do you see the implementation of ethically guidelines?*

	M (STD)
Digital Twins are in fact augmenting human capabilities. That implies that humans decide at the end guided by algorithms, medical professionals and doctors GP's.	*2.22 (1.99)*
A medical professional should always be in the loop and a Digital Twin-based solution is only a tool for a medical professional to aid in the process of decision making	*3.33 (1.76)*
Complete transparency on the how and why we use a Digital Twin should give trust to the patient to use a Digital Twin	*3.44 (1.57)*

4.6 Implementation list Top 3 rated according to importance

***Table 9**: Rated top 3 on the Question, 'How do you see the implementation of ethical guidelines? Likert scale (1- Strongly disagree – 7- Strongly agree)*

	M(STD)
Digital Twins are in fact augmenting human capabilities. That implies humans decide at the end	*6.33 (1.56)*
Complete transparency on how and why we use a Digital Twin should give trust to the patient to use a Digital Twin.	*6.22 (0.79)*
A medical professional should always be in the loop. A Digital Twin-based solution is only a tool for a medical professional to aid in the process of decision making.	*6.11 (1.73)*

The notion that Digital Twins augment human beings implies that humans/patients ultimately are the ones making decisions regarding which routes to traverse in a given medical scenario. With this notion of augmenting people's lives, experts agreed that a complete transparency is needed as to the how and why decisions are made so that a Digital Twin could be trusted by patients. A quote on this matter from one of the experts '1/ AI at the service of the individual, where it should help in optimizing healthcare for that specific person, based on the available data of that individual mixed with general data available from similar cases/persons; and 2/ AI at the service of society, where all data is anonymized and helps to finetune or optimize advice or support for several individuals, stakeholders or for society overall. This difference in data collection and usage should be made explicit from the start. The individual should have the decision power in which personal data and/or context is introduced, should be able to simulate how different versions of that data has an impact on the outcome (in order to finetune his/her input), and should have the decision power how the advice is being implemented in his/her own personal case (in line with how a patient decides to follow or not the advice from his/her doctor)'.

Research shows that ethics in Digital Twins combined with AI applications are extremely important especially in a healthcare environment; not only the privacy of the patient should be protected, but patients should always give their consent in using the data. Furthermore, the robustness of the technology should get great attention in terms of hacks and viruses. Guided by a medical professional to help and guide the patient towards the best outcome for his or her treatment (Barricelli, 2019), (Rasheed, 2020), (Bruynseels, 2018), (AI HLEG, 2019).

5. DISCUSSION

With this paper we set out to obtain a consensus view among experts regarding the materialisation, expectation, and the implementation of Digital Twins in healthcare. We worked together with experts who guide and stimulate start-ups in digital healthcare in the Belgian health industry and, more specifically, the futurist and healthcare start-up scene.
We choose to look at the definition of a Digital Twin where we adopt the definitions of Grieves, more specifically at a Digital Twin model where the data flows automatically between a physical object and a digital object in two directions. In such a combination, the digital object might also act as controlling instance of the physical object. Furthermore, we also looked at the 'lightweight version of a virtual model' as an enabler for isolated clinical trials or precision medicine.
Digital Twins in healthcare are still in early stage, yet it's a technology that moves fast in certain areas of healthcare. In our findings we can categorize them as important for the healthcare industry, the patients and society as a whole. As experts in this study are active in a diverse range of different aspects of digital transformation in healthcare, a wider perspective on the notion of a Digital Twin has been accessed. This is extremely important as there are a variety of different stakeholders in a Digital Twin process. These different stakeholders will need to collaborate and align their efforts to realize a Digital Twin. This multidisciplinary collaboration is equally important if not more than the technical challenges that are there (Fuller, 2020).

The Delphi study helped to reach a consensus quite rapidly. When experts ranked their input on the materialisation and implementation, the ranking agreement was low but significant. However, in a second round, when rating the top 3 there was a high agreement on the top 3 topics. As mentioned, each expert considering the materialisation, expectation, and implementation of Digital Twin with respect to their own domain, they held very diverse views and ideas about it. Another explanation could be that for current implementations of Digital Twins are premature and yet far from a complete implementation.

The materialisation of a Digital Twin is dependent on data sets that are generated by the healthcare industry in the broad sense of this term. This means that private companies gather data through commercialization of devices or mobile applications of all sorts to measure the health status of the individuals as individuals gather that information either voluntarily or pushed (Lupton D. , 2014). Through IoT services individuals also measure certain aspects of their environment and give way to the smart home. Within all this data, standards and interoperability of all these silos become extremely important as one also needs to create context around the data to fulfill the promise of accurate support decision making process. (Malakuti, 2018). In this study experts agreed on three main constructions: first, Digital Twin as a facilitator towards preventive healthcare, this finding is a common given in the literature on Digital Twins in healthcare in general (Fuller, 2020) (Rasheed, 2020) (Rivera, 2019). Second, experts see applications in Intensive Care Unit (ICU), a controlled lab environment where medical professionals can follow patients more accurately following certain parameters linked to any given patient. As a last, the construction of trial and error, where a Digital Twin could be operating as a test environment instead of operating and doing trial and error directly on patients, this notion leads way to personalised medicine and human centred patient care, where personalised medicine is tailoring medical treatments based on biomarkers and genetic information of a patient, while centred patient care is about a holistic view of a patient's lifestyle, 'considering the individual behind the patient' (El-Alti, 2019).

As mentioned, earlier healthcare applications entering the market are already making the first steps towards Digital Twins, as for example, Fibircheck and Epicare@home. Experts agreed on their expectation that through a trial-and-error process, a Digital Twin could give more insight in the remedial process of a certain pathology with a given patient. Digital Twins can give information on the past using historical data, the state of now as they are living data objects and always up to date. Moreover, through deep learning algorithms Digital Twins can predict or forecast a given situation for an individual. In addition, experts also agreed that given the ability of deep learning algorithms, Digital Twins can create a context around the data to improve upon the accuracy of health assessment and support decision making.

Experts also had high expectations regarding the application of a Digital Twin at a macro level, for example, to give an overview of the health status of a population - society as whole of a certain neighbourhood, a region, or a nation. In our introduction we mentioned the dangers of creating new digital divides, an elite class would benefit from Digital Twins. This emphasizes the need to 'democratize technologies', creating policies and regulations that are inclusive rather than being exclusive. (Rasheed,

2020) (Fukuyama, 2002).
Our panel agreed that 'Digital Twins should be seen as augmenting human beings, with humans/patients ultimately determining which decisions are made within the process. The autonomy and agency of patients in using a Digital Twin is an important aspect in the implementation of Digital Twins in healthcare (Bartneck, 2020). This implies that there should always be a human in the loop. 'A Digital Twin-based solution is only a tool for aiding a medical professional in their decision making' notes one of the experts. Experts argued for, 'complete transparency on how and why we use a Digital Twin should give trust to the patient to use a Digital Twin'.
When considering all the input given by the experts for this paper, one can conclude that Digital Twins hold some valuable promise to improve prevention and personalised medicine, providing more accurate and detailed views on a given individual/patient to support more accurate decision making in complete transparency, in collaboration and guided by a medical professional.

6. LIMITATIONS

This study was conducted in Belgium and the experts identified are representing views relevant in the local health system and industrial context. The questions we raised may be answered differently in countries with a more dominant private sector in healthcare, or with less stringent privacy regulations than the European Union. Further, the representation of industry in our study lacked the representation of pioneering companies, like Fibricheck or Byteflies mentioned above, or world leading industries in healthcare technology which can have a radically different perspective as to what is feasible or desirable.

Regarding the Delphi method as such, it was successful in reaching a consensus view. While experts disagreed at the first round, consensus could be reached within 2 of the 3 iterations. One of the reasons for the divergent views could be that the experts depart from their own domain of expertise and consider the needs of different sectors.

Future research should expand on topics that this study did not examine in depth, such as the ethical aspects of Digital Twins and the user (non-expert) perspective on these developments.

7. CONCLUSION

In this study we used the Delphi research method to get a broader view and consensus on the prospects of Digital Twins in healthcare. The main findings and conclusions of this study were that experts view a Digital Twin as a facilitator of preventive healthcare, a testing environment in an ICU controlled environment, where certain parameters of a patient are tracked, resulting in more accurately follow up on a given patient. The controlled environment also gives the ability to have a trial-and-error environment, that leads to personalised medicine and/or centred patient care. In addition, the expectations of a Digital Twin are seen as a way to get insight in remedial medical processes, creating context around the data that is available and giving oversight on the health status of a given population. Furthermore, experts see Digital Twins as a tool supporting decision making by medical professionals in complete transparency. Further research is needed on user perspectives and ethical aspects of Digital Twins in healthcare.

8. REFERENCES

AI HLEG, (2019). Ethics Guidelines For Trustworthy AI.https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

Bagaria, N. (2020). Health 4.0: Digital Twins for Health and Well-Being. pp. 143-152.

Barrett, M. A. (2013, September). Big Data and Disease Prevention: From Quantified Self to Quantified Communities. Big Data, pp. 168-175.

Barricelli, B. R. (2019). A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. IEEE Xplore, 7, pp. 167653-167671.

Bartneck, C. (2020). An Introduction to Ethics in Robotics and AI. SpringerBriefs in Ethics.

Bruynseels, K. (2018). Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm. Frontiers in Genetics, 31.

Byteflies.(n.d.).Retrieved from https://www.byteflies.com/

De Maeyer, C. (2020). Are Digital Twins becoming our Personal (predictive) Advisors? 'Our Digital Mirror of Who We Were, Who We Are and Who We Will Become'. HCI International 2020. 12208. Copenhagen: Gao Q., Zhou J. (eds) Human Aspects of IT for the Aged Population. Healthy and Active Aging. HCII 2020. Lecture Notes in Computer Science. Springer, Cham.

El-Alti, L. (2019). Person Centered Care and Personalized Medicine: Irreconcilable Opposites or Potential Companions? Health Care Anal, pp. 45-59.

Elayan, H. (2021). Digital twin for intelligent context-aware iot healthcare systems. Ieee Internet of Things Journal.

Etzkowitz, H. (2017). The Triple Helix: University–Industry–Government Innovation and Entrepreneurship (2nd ed.). Routledge.

Fibricheck.(n.d.).Retrieved from https://www.fibricheck.com/

Fukuyama, F. (2002). Our Posthuman Future: Consequences of the Biotechnology Revolution. Published by Farrar, Straus and Giroux. Macmillan.

Fuller, A. F. (2020). Digital Twin: Enabling Technology, Challenges and Open Research. IEEE Access, 8, pp. 108952-108971.

Gordon, T. J. (1994). The Delphi Method.

Grieves, M. (2014, January 3). Digital Twin: Manufacturing Excellence through Virtual Factory Replication. USA.

Grieves, M. (2017). Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. Kahlen FJ., Flumerfelt S., Alves A. (eds) Transdisciplinary Perspectives on Complex Systems.

Healthouse, Retrieved from https://www.health-house.be/en/

Helmer, O. (1950-60). Retrieved from https://www.rand.org/

Justice, A. (n.d.). Algoritmic Justice League: Unmasking AI harms and Biases. Retrieved from https://www.ajl.org/

Kritzinger, W. (2018). Digital Twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine, 51(11), pp. 1016-1022.

Lakki, H. (2019). Prototyping a Digital Twin for Real Time Remote Control Over Mobile Networks: Application of Remote Surgery. IEEE Access.

Lewis, D. (2021). Contact-Tracing Apps Help to Reduce Covid Infections. Nature.

Lupton, D. (2014). Self-Tracking Modes: Reflexive Self-Monitoring and Data Practices. SSRN Electronic Journal.

Lupton, D. (2019). Data selves, More-than-Human Perspectives. Wiley.

Malakuti, S. (2018). Architectural aspects of digital twins in IIoT systems. Proceedings of the 12th European Conference on Software Architecture: Companion Proceedings ECSA '18, Association for Computing Machinery, N (pp. 1-2). ACM Digital Library.

Mullen, P.M., (2003). Delphi:myths and reality, Journal of Health Organization and Management, Vol 17 No1, pp 37-52

O' Neill, C. (2016). Weapons of math destruction. Crown Books.

Persistent. (2020). Digital Twin Whitepaper. Retrieved from https://www.persistent.com/wp-content/uploads/2020/01/digital-twins-whitepaper.pdf

Rasheed, A. (2020). Digital Twin: Values, Challenges and Enablers From a Modeling Perspective. IEEE Access, 21980-22012.

Renzi, A.B. (2015). Delphi method to explore future scenario possibilities on technology and HCI. Marcus A. (eds) Design, User Experience, and Usability: Design Discourse. Lecture Notes in Computer Science. Springer, Cham.

Rivera, L. F. (2019). Towards continuous monitoring in personalized healthcare through digital twins. Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering (CASCON '19) (pp. 329-335). IBM Corp.

Rodríguez, P. (2021). A population-based controlled experiment assessing the epidemiological impact of digital contact tracing. Nat Commun 12, 587.

Saaty, T. (2008). Decision making with the analytic hierarchy process . Int. J. Services Sciences, Vol. 1, No. 1.

Saddik, A. (2018). Digital Twins: The Convergence of Multimedia Technologies. IEEE MultiMedia,, 87-92.

Searl, M. (2010). It is time to talk about people: a human-centered healthcare system. Health Res Policy Syst.

Smith, G. (2016). Surveillance, Data and Embodiment: On the Work of Being Watched . Body & Society, Vol.22, 108-139.

Schwartz, (2020). Digital Twins and the Emerging Science of Self: Implications for Digital Health Experience Design and “Small” Data. Frontiers in Computer Science, 31.

World Economic Forum. (2020, April 7). Retrieved from What can smart cities tell us about covid-19: https://www.weforum.org/agenda/2020/04/smart-cities-technology-coronavirus-covid19/

Yang, C. (2019). Exploring Design Guidelines of Using User-Centered Design in Gamification Development: A Delphi Study. International Journal of Human–Computer Interaction, 1170-1181.

A Proposed Framework for Developing User-Centred Mobile Healthcare Applications for the Biggest Annual Mass Gathering (Hajj) Post COVID-19

Nowman Islam
School of
Computing and Engineering
University of West London
St. Mary's Road, Ealing, UK

21339907@student.uwl.ac.uk

Aamir Anwar
School of
Computing & Engineering
University of West London
St. Mary's Road, Ealing, UK

21452391@student.uwl.ac.uk

Ikram Rehman
School of
Computing & Engineering
University of West London
St. Mary's Road, Ealing, UK

Ikram.Rehman@uwl.ac.uk

Abstract – The Hajj pilgrimage being the largest annual mass gathering globally with two to three million participants from over 180 counties, will remain a high priority for diseases surveillance for future epidemics or any other international public health emergencies with rapid scalability. This paper highlights the importance of monitoring mass gatherings during a pandemic and how mHealth applications can reduce the burden on health facilities during a mass gathering and tackle future infectious diseases outbreaks. The paper also highlights the importance of developing a user-centred application when designing for a diverse group of users with a shared purpose. As a result, a framework has been proposed to update the current applications or design and develop future mobile health applications. The framework has been developed based on the rationale and evidence found in the literature.

KEYWORDS – COVID-19, MASS GATHERING, HAJJ, MHEALTH (MOBILE HEALTHCARE), PERSUASIVE SYSTEM DESIGN (PSD), SENTIMENT ANALYSIS, USER-CENTRED DESIGN (UCD), DOUBLE DIAMOND.

1. Introduction

According to the World Health Organisation (WHO) a "mass gathering" is any gathering of people in a location where the health systems and facilities are pushed and strained to their boundaries or more [1].

There are currently over 1.8 billion Muslims worldwide, making "Islam" the second-largest and fastest-growing religion globally, as reported by Pew Research Centre [2]. Hajj is one of the five pillars of Islam, and it consists of visiting the holy city of Mecca, which is located in the Kingdom of Saudi Arabia (KSA). In an editorial article on the Bulleting, WHO stated that the Hajj pilgrimage is the largest annual mass gathering in the world, with 2 to 3 million adult Muslim participants from over 180 countries in the KSA [3].

In January 2020, WHO announced a Public Health Emergency of International Concern (PHEIC), identifying the Covid-19 as a worldwide pandemic. It is not the first time that a deadly virus has speeded around the world. We have previously seen SARS (Severe acute respiratory syndrome) and MERS (Middle East respiratory syndrome). However, this is the first time that a virus had a global effect [4]. A covid-19 mortality analysis by Johns Hopkins University & Medicine shows that nearly every county on the planet has been affected by the deadly virus [5].

This pandemic means that not only are the personal lives affected by the disease but entire economies and nations are also affected, including KSA, where the biggest annual mass gathering takes place. As a result, one of the most critical questions that had significant attention in the Muslim world is whether the KSA authorities would allow the Hajj to take place on a full scale, reduce scale, or temporarily discontinue the entire pilgrimage, which originated in 2000 B.C. [6].

Due to the rapid community transmission of the deadly virus, the authorities responded to the dire state by scaling down the Hajj pilgrimage to only 1000 participants residing in the Kingdom compared to the 2.5 million participants in the previous year [7].

Digital technologies advancement in the health sector, including Big data and artificial intelligence (AI), have supported and facilitated many countries around the globe to stop the spread of the deadly Covid-19 infections [8]. In a digital health laws and regulations of Saudi Arabia report of 2021 [9], a range of mobile health applications for the KSA is recommended. These applications are developed by leading authorities such as the Ministry of Health.

It is uncertain if the KSA authorities' current mobile healthcare systems are ready to support the existing healthcare facilities of the Hajj pilgrimage for a post-Covid-19 scenario with a full-scale pilgrimage. This is important because participants are from over 180 counties with different socio-economic backgrounds, technology capabilities and language barriers.

The main contribution of this paper is a proposal of a framework to the Saudi Ministry of Health (MOH) for designing and developing future Mobile health applications or updating the current applications to be more human-centred with an increased user

experience level and adoption rate based on the evidence gathered from the literature dating from 2012 to mostly 2021 on topic such as “Mass gathering health threats”, “Hajj health issues”, “Mobile Apps during Hajj”, “Saudi mHealth apps”,“User review analysis”, “persuasive mHealth” and “Design Thinking in mHealth”.

2. Related Work

2.1 Types of Mass Gatherings and their health threats

In a systematic review of public health threats in mass gatherings published by Cambridge University Press in June 2019 [10], it has been mentioned that there are mainly three types of mass gatherings (MGs) taking place around the globe and they are: 1) religious MGs, 2) festival MGs and 3) sporting MGs. The review aimed to identify and assess the health threats associated with each type of the MGs mentioned above. The study was performed on 45 articles out of 1264 references found through the Web of Science, Medline, and Scopus database. The main findings of the review are as follows:

1. In the festival-related mass gatherings, the most common health risk were alcohol and drugs-related issues.
2. In the sporting type of mass gatherings, the most common health risk identified is injuries. Alcohol, drugs, and infectious diseases are other risks identified.
3. In the religious type of mass gathering, the most common health problem identified was infectious diseases. In addition, environment and road traffic-related risks were also found.

The review concluded by recommending that the organisers of the mass gatherings assess the type of the MGs and the health and safety needs of the participants with the proper planning for the healthcare facilities critically.

All the findings by the authors are confirmed in a report by WHO published in 2015 titled "Public Health for Mass Gathering: Key Considerations [11]. The report aims to provide advice, recommending methods and frameworks to mass gathering organisers on prevention, detection and management of public health incidents [10].

2.2 Health issues and current health system during the Hajj Pilgrimage

A research study has reviewed 60 articles on health issues during the hajj pilgrimage from 2005-2014 [12]. The authors primarily retrieved 335 research papers from PubMed, a leading and trusted biomedical and life sciences literature database. The paper's goal was to provide recommendations on the policies to prevent the pilgrims' health risks. The paper's outcome resulted in three categories using only qualitative analysis: health services, communicable disease, and non-communicable disease. The authors reported that respiratory-related diseases are most common during the pilgrimage. There is a low satisfaction rate on the physicians' services, and that special attention is needed to improve the hajj participants' experience. It is also reported that more evaluation studies are required with Saudi's health services during the Hajj period to make recommendations for improvements. The paper concluded by stating that the optimal utilisation of the current Hajj health system's collected data is yet to be achieved [12].

2.3 mHealth apps by the KSA authorities during Covid-19

A recent research paper published in 2020, "Digital response during the covid-19 pandemic in Saudi Arabia", highlighted that the KSA authorities partnered with the private sector companies to develop and launch approximately 19 platform/applications to support the public healthcare services during the Covid-19 pandemic. The authors did not mention if these applications are an excellent fit to serve the largest annual mass gathering hosted in the Kingdom for a post-Covid-19 scenario. On the contrary, the authors recommended reducing the number of mHealth apps by merging the various features [13].

Another paper also published during Covid-19 in 2020 [13] evaluated the mHealth applications developed by the Ministry of Health (MOH) for public use. The study aimed at identifying and understanding the design of two applications in particular by evaluating their persuasive features. Both the applications included in the study were developed by the MOH. The first is called "Sehha", a telemedicine application where users can remotely consult via text, voice message, images and videos with qualified doctors approved by MOH. The second app is called "Mawid", which enables patients to book their appointments, cancel or reschedule with health centres and hospitals. The author recommended improving the design and including persuasive features in the application published by MOH.

The lack of understanding of the human context by the app developers and other stakeholders can result in developing and launching products that will remain primarily unused and not meeting the project objectives. The design thinking principles is a robust methodology that can close the gap by involving the user needs and feedback throughout the development of new healthcare systems, especially mobile healthcare applications [14].

As mHealth applications are becoming a popular way to motivate users toward healthy behaviour, the design and inclusion of the functionalities play a significant role in persuading the public to use those apps. Therefore, the impact of mHealth apps is dependent on their adaptability to be acceptable to various users [13].

An industry report titled "mHealth 2030" by ORCHA, the world's leading health app evaluation and advisor organisation [15], discussed that users would be the principal stakeholder for the failure of success of the mHealth ventures. In the same report, an expert comment explains that there are and will be a surplus number of mHealth apps for the users to choose from, and not all of them are developed to keep the user's interest. Therefore, only the applications that meet the user need will succeed in the longer run by carefully identifying and including the unique selling points (USP), making regular updates and most importantly, integrating persuasive design as we see in the gaming industry [15].

3. proposed framework

Our proposed framework to design, develop new mHealth applications for the Hajj pilgrimage consists of using sentiment analysis for detecting the user experience state on the current apps developed by KSA and the Persuasive System Design (PSD) framework within the British Design Council's Double Dimond methodology to influence, reinforce or change health-related behaviours or attitudes of users during the Hajj pilgrimage.

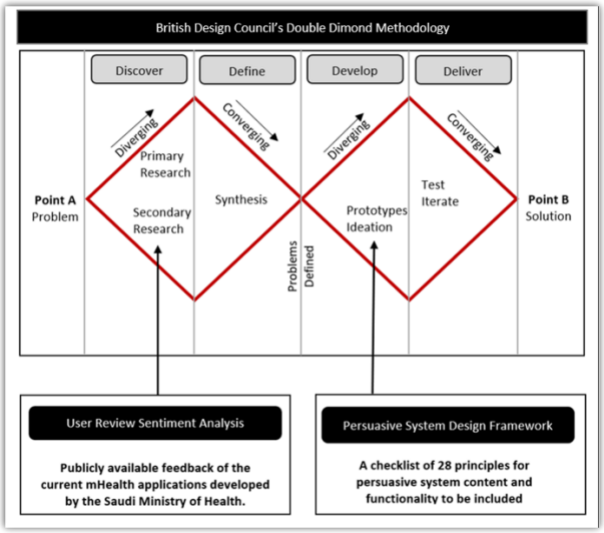


Figure 1: *Proposed Framework*

3.1 British Design Council's Double Dimond

The Double Dimond method works in a way that both designers and non-designers can understand the processes. They represent a process of exploring the problems with a broad view called "diverging thinking" and then taking focused actions called "convergent thinking" [16]. The reason for selecting this methodology in the framework is the fundamental characteristic of its human-centred approach. This is a collaborative way between designers and other stakeholders such as users in a co-creation or co-participatory setting, shifting the attitude from designing "for users," which is the traditional app development process, to a human-centred approach of designing "with users"[14].

There are mainly four phases in this methodology:

1. **Discover:** In this phase, designers need to address the problem and understand them clearly. In the context of the Hajj, a question could be "how do I design and develop an app that tracks pilgrims, preventive measures and monitors diseases for users with different socio-economic backgrounds, languages and technology capabilities". Authorities should recruit diverse teams of experienced pilgrims from various countries when carrying primary research through surveys, observations, and user diaries. This is because experienced pilgrims know the Hajj rituals at best, and they are also familiar with the locations and the health implications. In the secondary research, the authorities should extensively study the Hajj statistics yearbook, where a detailed breakdown of the pilgrims' demographics is given. Other relevant secondary research can be carried out once the participants' countries are identified.

2. **Define:** This phase is where all the unstructured research findings get refined through in-depth research tools such as focus groups and user journey mappings.

3. **Develop:** This phase is where ideas are brainstormed on clearly defined problems from the previous phase. It is suggested to seek inspiration from successful mHealth applications from other countries and co-design with the recruited people from the earlier phases.

4. **Deliver:** This phase is where small scale solutions are developed and tested out with a set of users for collecting essential data such as user acceptance rate etc. This stage is highly iterative, so it is not a linear process.

3.2 User Reviews Sentiment Analysis

Sentiment Analysis (SA) is a data mining technique used to analyse social media, political campaign, or product reviews data, to name just a few. This is a key indicator of user satisfaction and user experience [17]. SA is a process that uses artificial intelligence (AI) to assign a negative (-1) value or a positive value (+1) to a piece of text. An example would be "the app is well designed, but it keeps crashing". SA has the potential to analyse and filter out subjective opinions of users from an app review section. The most crucial benefit of this technology is the ability to find out the user sentiments through subjective words such as "awesome", "great",

"lifesaver", "useless", "crushing", "confusing", "pointless", and so on. [18].

We propose to embed SA in the double diamond's discovery phase along with other designer tools where the design team performs primary and secondary research. Therefore, analysing the publicly available user reviews of the existing apps developed by the KSA can give valuable insights into the current state of the user experience.

3.3 Persuasive Systems Design Framework (PSD)

The PSD is a framework developed to design and evaluate persuasive systems [19]. For an application to be persuasive, it is vital to analyse the use context of the system. In this case, it is the Hajj pilgrimage. It is also essential to understand the role of the persuader (Ministry of Health) and persuadee (Hajj participants), their communication channel, and what content is being communicated. Moreover, the design and the innovation team should analyse users' needs, motivations, technical abilities, personalities, and interest, among many other factors [19].

The PSD framework consists of a checklist of 28 design principles in four categories based on activities a user is set to accomplish, as shown in Table 1

Table 1: *PSD Design Principles Checklist* [19].

Primary Task Support	Dialogue Support	Social Support	System Credibility Support
Reduction	Praise	Social learning	Trustworthiness
Tunnelling	Rewards	Social comparison	Expertise
Tailoring	Reminders	Normative influence	Surface credibility
Personalization	Suggestion	Social facilitation	Real-world feel
Self-monitoring	Similarity	Cooperation	Authority
Simulation	Liking	Competition	Third-party endorsement
Rehearsal	Social role	Recognition	Verifiability

As discussed in the related work section, WHO highly endorses the human-centred approach when designing mobile healthcare applications [20].

In Figure 2. we can see that the Persuasive System Design has been embedded within the British council's double diamond's third phase (develop). As illustrated in Table 1, the inclusion of any relevant principles will be beneficial when brainstorming ideas for prototyping mHealth applications [21]. As an example, a Minimum Viable Product (MVP) would adhere the trustworthiness, and authority principles from the Table 1. by including truthful information by authoritative figures such as Ministry of Health or WHO.

CONCLUSION

This paper presents the preliminary findings from the literature on the importance of mass gatherings during a pandemic and how mHealth applications help the health facilities in mass gatherings. We also discussed the importance of developing a user-centred application when designing for a diverse group of users with a shared purpose. Findings included the endorsement of a user-centred approach for mHealth application from the WHO and research papers in prestigious journals. This paper proposed a new theoretical framework for the KSA authorities to facilitate the design and development of future mHealth applications. The framework has the potential to improve the user acceptance rate and engagement level, which currently lacks in mHealth apps in the context of the Hajj pilgrimage.

4 REFERENCES

[1] "Coronavirus disease (COVID-19): Mass gatherings." https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19-mass-gatherings (accessed Apr. 23, 2021).

[2] "Muslims and Islam: Key findings in the U.S. and around the world | Pew Research Center." https://www.pewresearch.org/fact-tank/2017/08/09/muslims-and-islam-key-findings-in-the-u-s-and-around-the-world/ (accessed Apr. 28, 2021).

[3] Z. A. Memish and A. A. Al-Rabeeah, "Public health management of mass gatherings: The saudi arabian experience with MERS-CoV," *Bulletin of the World Health Organization*, vol. 91, no. 12. World Health Organization, p. 899, 2013, doi: 10.2471/BLT.13.132266.

[4] N. N. Chathappady House, S. Palissery, and H. Sebastian, "Corona Viruses: A Review on SARS, MERS and COVID-19," *Microbiol. Insights*, vol. 14, p. 117863612110024, Jan. 2021, doi: 10.1177/11786361211002481.

[5] "Mortality Analyses - Johns Hopkins Coronavirus Resource Center." https://coronavirus.jhu.edu/data/mortality (accessed Apr. 29, 2021).

[6] "A Brief History Of the Hajj - TIME." http://content.time.com/time/world/article/0,8599,1864624,00.html (accessed Apr. 28,

2021).

[7] "Hajj 2020: What you need to know about this year's pilgrimage | Coronavirus pandemic News | Al Jazeera." https://www.aljazeera.com/news/2020/7/25/hajj-2020-what-you-need-to-know-about-this-years-pilgrimage (accessed Apr. 29, 2021).

[8] S. Whitelaw, M. A. Mamas, E. Topol, and H. G. C. Van Spall, "Applications of digital technology in COVID-19 pandemic planning and response," *The Lancet Digital Health*, vol. 2, no. 8. Elsevier Ltd, pp. e435–e440, Aug. 01, 2020, doi: 10.1016/S2589-7500(20)30142-4.

[9] "Digital Health 2021 | Saudi Arabia | ICLG." https://iclg.com/practice-areas/digital-health-laws-and-regulations/saudi-arabia (accessed May 07, 2021).

[10] M. Karami *et al.*, "Public Health Threats in Mass Gatherings: A Systematic Review," *Disaster Medicine and Public Health Preparedness*, vol. 13, no. 5–6. Cambridge University Press, pp. 1035–1046, Dec. 01, 2019, doi: 10.1017/dmp.2018.161.

[11] "WHO | Public Health for Mass Gatherings: Key Considerations," World Health Organization, 2015. Accessed: May 06, 2021. [Online]. Available: http://www.who.int/ihr/publications/WHO_HSE_GCR_2015.5/en/.

[12] M. R. Aldossari, A. Aljoudi, and D. Celentano, "Health issues in the hajj pilgrimage: A literature review," *Eastern Mediterranean Health Journal*, vol. 25, no. 10. World Health Organization, pp. 744–753, Oct. 01, 2019, doi: 10.26719/2019.25.10.744.

[13] A. Shati, "Mhealth Applications Developed by the Ministry of Health for Public Users in KSA: A Persuasive Systems Design Evaluation," *Heal. Informatics - An Int. J.*, vol. 9, no. 1, pp. 1–13, 2020, doi: 10.5121/hiij.2020.9101.

[14] K. Tschimmel, "Design Thinking as an effective Toolkit," *Action Innov. Innov. from Exp. Barcelona*, pp. 1–20, 2012.

[15] "mHealth 2030 Report - ORCHA." https://orchahealth.com/mhealth-2030-report/ (accessed Jun. 01, 2021).

[16] "What is the framework for innovation? Design Council's evolved Double Diamond | Design Council." https://www.designcouncil.org.uk/news-opinion/what-framework-innovation-design-councils-evolved-double-diamond (accessed May 09, 2021).

[17] "Sentiment Analysis turns Customer Reviews into Insights - Vendasta Blog." https://www.vendasta.com/blog/sentiment-analysis-turns-reviews-into-insights/ (accessed May 09, 2021).

[18] M. Hassounah, H. Raheel, and M. Alhefzi, "Digital response during the COVID-19 pandemic in Saudi Arabia," *Journal of Medical Internet Research*, vol. 22, no. 9. JMIR Publications Inc., p. e19338, Sep. 01, 2020, doi: 10.2196/19338.

[19] I. Adaji and J. Vassileva, "Evaluating Persuasive Systems Using the PSD Framework," *undefined*, 2016.

[20] WHO Global Observatory for eHealth., *MHealth : new horizons for health through mobile technologies.* Switzerland: World Health Organization, 2011.

[21] H. Oinas-Kukkonen and M. Harjumaa, "Persuasive Systems Design: Key Issues, Process Model, and System Features," *Commun. Assoc. Inf. Syst.*, vol. 24, pp. 485–500, 2009, doi: 10.17705/1CAIS.02428.

Developing a clinical decision support system for paediatric shoulder instability

Amina Yaqub
Keele University
School of Allied Health Professions/University Hospitals of Leicester
amina.yaqub@hotmail.co.uk

Michael McCluskey
University of Liverpool
School of Health Sciences
m.mccluskey@liverpool.ac.uk

Edward Chadwick
University of Aberdeen
School of Engineering
edward.chadwick@abdn.ac.uk

Fraser Philp
University of Liverpool/
School of Health Sciences
f.philp@liverpool.ac.uk

Paediatric shoulder instability is a complex condition which may benefit from clinical decision support systems. The study aims were to elicit the types of information used in clinical decision making, and get early stakeholder involvement to inform the design of a decision support system. Knowledge elicitation was achieved using an online questionnaire distributed to registered physiotherapists. A wide range of factors used in clinical decision-making process were identified. These may inform the development of a clinical decision support system in conjunction with clinicians self-reported knowledge regarding 3D motion analysis and its perceived. This early stakeholder involvement is important for ensuring that subsequently developed systems are clinically applicable.

User centre design; knowledge elicitation; stakeholder involvement; clinical decision support systems; paediatric shoulder instability; rehabilitation; physiotherapy; clinical decision making; movement analysis.

1. INTRODUCTION

Shoulder instability (SI) that is, complete or partial dislocation of the shoulder, is common in children (overall incidence of 23.9 to 164.4 /100,000-person years in 10 to 16-year olds, (Leroux et al., 2015)). Current methods for diagnosing and treating SI in children are inadequate (Philp et al., 2021). Time to diagnosis can be delayed by up to two years and failure rates with rehabilitation are high with 70% to 90% of children continuing to suffer recurrent instability within 1 to 2 years after their initial instability episode (Longo et al., 2016). SI can result in pain, decreased movement and disability. Recurrent SI can damage the shoulder joint resulting in early arthritis onset (odds ratio 19.3) (Deitch et al., 2003, Marx et al., 2002).

1.1 CONCEPT OF A TOOL FOR CLINICAL DECISION MAKING IN PAEDIATRIC SHOULDER INSTABILITY

Mechanisms for SI are not well understood and clinical decision making is complex. Clinical Decision Support Systems (CDSS) may help in reducing time to diagnosis and improve patient outcomes. CDSS involving, three-dimensional (3D) motion analysis (figure 1) and clinical algorithms for imaging selection have been shown to improve diagnostic accuracy in SI and other upper-limb conditions (Moroder et al., 2020, Brunner et al., 2020). Despite their utility, they are not widely used in practice. A possible reason for this is the lack of early end-user involvement and explicit mapping of information used in clinical decision-making.

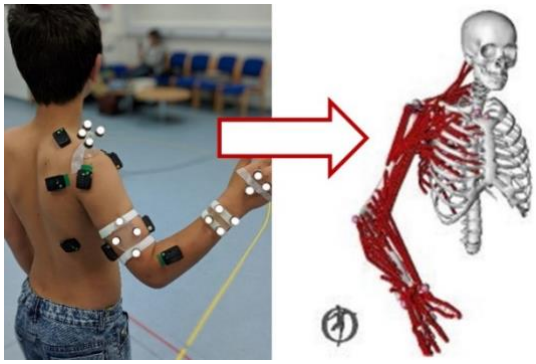


Figure 1: *3D motion analysis of a child, data from which is then inputted into a musculoskeletal OpenSim model*

Early stakeholder involvement is imperative for ensuring a CDSS is clinically applicable (Lane et al., 2016). This is particularly important for complex conditions, such as SI, where there are potentially significant amounts of information that may be derived from multiple sources e.g. CDSS involving 3D-movement analysis. It is also unclear what information would be useful in clinical decision making. There is a risk that the user will be overwhelmed with information and interfaces which are complex to navigate or present information that is redundant or difficult to interpret, will limit engagement (Jaspers et al., 2011). Performance of a CDSS is also dependent on the availability of appropriate patient characteristics and the systems knowledge base. In paediatric shoulder instability, it is not clear which patient characteristics should be included and if there is a sufficient knowledge base for the system to refer to.

The aims of this study are therefore to 1) elicit the types of information used for clinical decision making in paediatric shoulder instability and 2) get early

stakeholder involvement prior to development, to inform the design of a decision support system, which may integrate 3D movement analysis data.

2. METHODS

Ethical approval was gained from the School of Allied Health Professions Student Project Ethics Committee, Keele University. Knowledge elicitation was carried out using a web-based questionnaire (at https://doi.org/10.5281/zenodo.5018670). The questionnaire was developed by the clinical and non-clinical authors, informed by the literature and evaluated for face validity prior to recruitment. Participants were recruited over a two-month period, from November 2019 to January 2020, through social media platforms and special interest/ professional group mailing lists. Participants were included if they were qualified physiotherapists, currently working adult and paediatric SI patients. Participants were excluded if they were not registered physiotherapists or did not assess patients with SI.

3. RESULTS

A total of 19 responses (12F:7M) were received. Respondents had been qualified for an average of 14.6 years (SD 8.1) and worked primarily in orthopaedic settings. Eighteen respondents worked in the UK and one worked in Egypt. Only one participant had previous experience working in a movement analysis service.

3.1. Results for diagnostic information

Six participants did not report using any classification systems. The Stanmore classification/Bayley triangle (Lewis et al., 2004) framework the most frequently (n=11), followed by the Gerber and Nyffeler (Gerber and Nyffeler, 2002) (n=2), and the Thomas and Matsen (Thomas, 1989) frameworks (n=1). A summary of the subjective factors' clinicians found important for diagnosing paediatric shoulder instability, and frequency of selection, are reported in Table 1. The number of factors selected by any individual ranged from five to 16, with a mode of eight. No single subjective or objective factor was identified universally by all respondents. A summary of the objective factor's clinicians found important for diagnosing SI, and frequency of selection, are reported in Table 2. The number of factors selected as being important for shoulder instability by any individual ranged from two to 16, with a mode of seven.

Table 1: *Frequency of subjective factor selection*

Subjective factor	Number of responses
Previous episode(s) of SI	18
Sensation(s) of SI	17
Previous shoulder injury, Injury mechanism consistent with SI	16
Age, Previous surgery related to instability	15
Loss of Function	13
Pain, Weakness	11
Previous investigations	10
Gender	9
Previous treatment	8
History of repetitive micro trauma, Family history of SI	7
Locking, Deformity	3
Stiffness	2
Catching, Swelling, Clicking	1

Factors separated by (,)

Table 2: *Frequency of objective factor selection*

Objective factor	Number of responses
Apprehension relocation test	16
Sulcus Sign	14
AROM Apprehension	12
Beightons score greater than or equal to 4/9	11
Loss of rotation (internal & external), Posture e.g. asymmetry, scapular winging, atrophy, passive range of movement apprehension	10
Loss of active abduction, Posterior apprehension, Scapula Dyskinesis	9
Inferior apprehension	8
Anterior shift load, Loss of active flexion	6
Drawer tests, Loss of active adduction, Loss of active extension	5
Pain/stiffness with passive range of movement, posterior shift load	4
Tenderness on palpation	3
Release Test	2
Arc Test, Hyperabduction test, Jerk test, Shoulder Symptom Modification Procedure, Rotator cuff weakness	1

Factors separated by (,)

3.2. Results for rehabilitation protocols used

Overall seven protocols for rehabilitation were identified. Only two respondents reported not using any protocols. The Derby Shoulder Instability

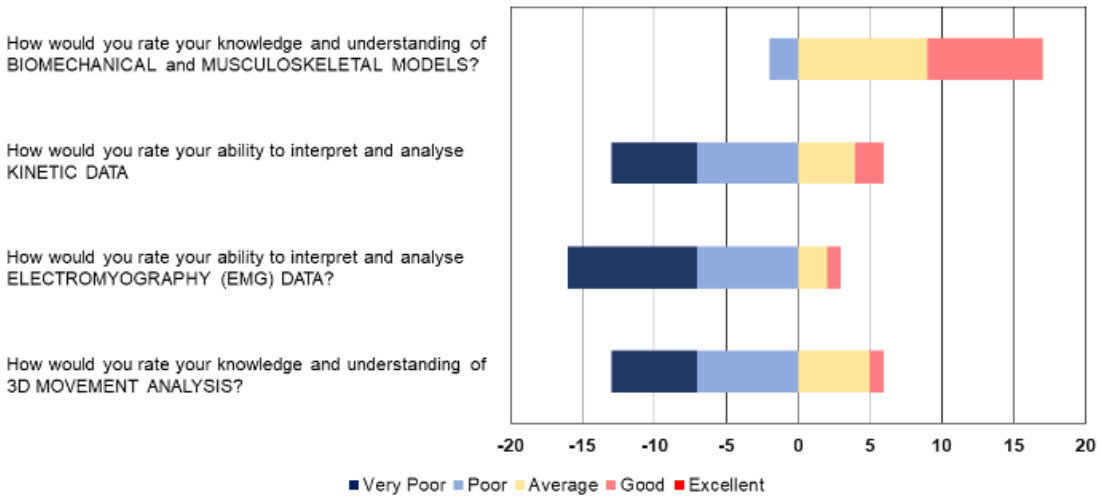
Rehabilitation Programme was the most frequently identified (Bateman et al., 2019) (n=9), followed by Consultant developed/led protocols (n=7), Department specific protocols (n=3), the Watson Multidirectional Instability Program (Watson et al., 2016, Watson et al., 2017) (n=2), Trust specific protocols (n=1), Individualised protocols (n=1) and protocols based on external courses (n=1).

3.3. Results for respondent's knowledge and understanding, and perceived usefulness of movement analysis

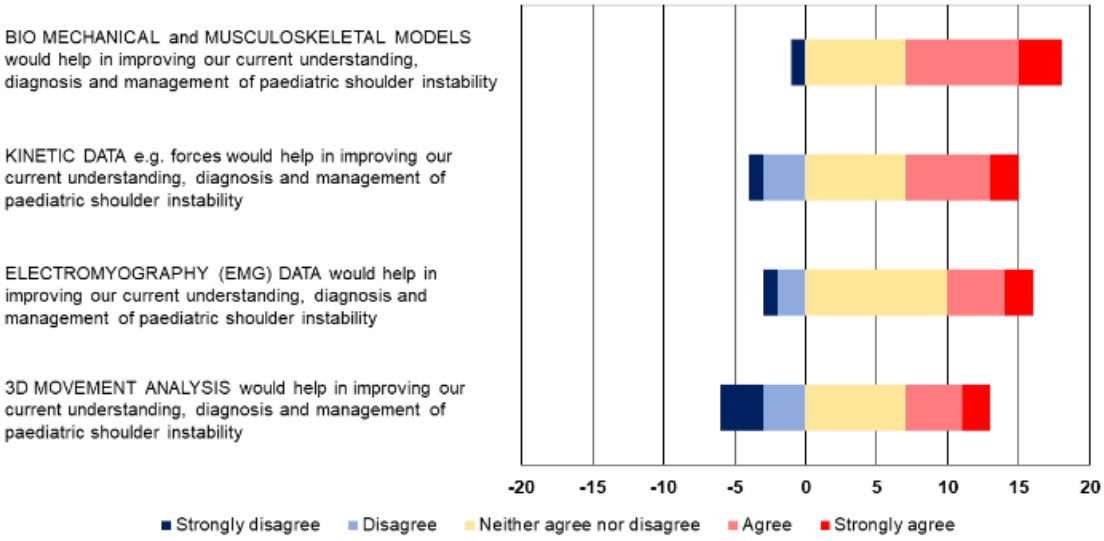
Overall results are presented in figure 5. Respondents reported having average (47.4%) or good knowledge (42.1%) of biomechanical and musculoskeletal models and having poor or very poor knowledge or ability (more than 60%) regarding 3D movement analysis and interpretation of electromyography (EMG) and kinetic data. The most frequently selected answer was that respondents neither agreed nor disagreed that kinetic data, EMG and 3D movement data would improve their current understanding, diagnosis and management of paediatric SI. The use of biomechanical and musculoskeletal models was viewed positively by the majority of respondents with 42.1% and 15.8% of respondents agreeing or strongly agreeing respectively.

Figure 5. *Respondents knowledge and understanding of movement analysis and perceived usefulness of motion analysis in improving understanding, assessment and management of paediatric shoulder instability*

A. Respondents knowledge and understanding of movement analysis



B. Respondents perceived usefulness of motion analysis in improving understanding, assessment and management



4. DISCUSSION

The aims of this study were to 1) elicit the types of information used for clinical decision making in paediatric shoulder instability and 2) get stakeholder involvement in the early stages of development, to inform the design of a decision support system, which may integrate 3D movement analysis data. Our results indicate that there is good rationale for the use of a CDSS which may ultimately reduce

practice variation and improve patient outcomes. We have also identified explicit factors that could be used for informing the initial development of a CDSS, associated interface and future work.

There was considerable variation in the number of selected subjective and objective factors. This is consistent with other published studies reporting between one to 16 factors (Kuhn et al., 2011). The most frequently selected subjective and objective factors were history or sensation of SI, and positions which made the patient report SI. It is not clear what the minimum number of criteria needed to accurately diagnose shoulder instability is, or which factors are considered the most important by individual therapists. This poses challenges for the development of a CDSS, as a large number of factors will increase imputation burden and limit clinical integration. Additionally, for cases in which clinicians use fewer factors, the accuracy of the CDSS will be negated if required factors not deemed important by individual clinicians are omitted. Further work is required to identify the agreed minimum number of factors required for accurate diagnosis. This may be achieved by using alternate methodologies such as Delphi and nominal focus group techniques.

Agreed classification systems, management pathways and suitable knowledge base are required for an effective CDSS. Existing systems are inadequate given their complexity and poor evidence base (Warby et al., 2018). Furthermore, they are prone to error as 1) there is poor agreement between patient reported and clinician derived factors (Kuhn et al., 2011) 2) “objective factors” e.g. joint integrity tests, lack sensitivity and specificity and are not always equivalent true physiological or biomechanical measures. This may explain why these systems have not been universally adopted e.g. no classification framework was the 2nd most frequent response (n=6).

It was identified that whilst motion analysis may enhance our understanding of paediatric shoulder instability, a lack of experience and knowledge regarding motion analysis may have affected clinicians perceived usefulness of these methods. Implementation of 3D motion analysis into routine clinical pathways has been done successfully in other domains (Laracca et al., 2014). Currently, however movement analysis requires referral into a specialist centres and so their integration into the clinical pathway requires further investigation. Respondents neither agreed nor disagreed about the use of 3D movement, kinetic and EMG data for the same purpose. These results are surprising given the interpretation of biomechanical/musculoskeletal models are dependent on 3D movement, kinetic and EMG data. A possible reason for this may be limited experience and knowledge of these measurement methods. If a CDSS is to incorporate the information derived from 3D motions analysis, future work or training may be needed to familiarise clinicians with these measurement methods and identify what additional information derived from motion analysis would be useful.

4.1. Limitations

We were unable to report a response rate as the questionnaire was distributed using social media platforms and special interest/ professional group mailing lists. Whilst not all physiotherapists manage patients with shoulder instability, it is recognised that our proportionate sample size is small given there are more than 50,000 registered physiotherapists in the UK (HCPC, 2017). Patients with SI may also be managed by other healthcare professionals and it would be beneficial to include all relevant stakeholders in the CDSS design process. Whilst a larger sample size may have highlighted additional factors, it is unlikely that these will be helpful given the considerable variation in practice already observed. Self-reported knowledge of clinicians against their actual performance was not verified and may account for our results. Despite a correlation between perceived and actual knowledge (Chan and Zang, 2007), healthcare professionals overestimate their actual knowledge and ability when less experienced and operating in new domains (Caspi et al., 2006). Further work is therefore required to establish the actual knowledge and ability of clinicians to interpret information derived from motion analysis, if this is to be integrated in a CDSS.

5. CONCLUSION

There are considerable variations in practice, for diagnosis and management of assessment, and this may be addressed by the development of an appropriate system. Stakeholder involvement identified that overall respondents agreed that biomechanical and musculoskeletal models would help to improve assessment and management of paediatric shoulder instability. It is recognised that prior to development of a clinically applicable systems, further work is needed to develop consensus and minimally required criteria for the assessment and diagnosis of paediatric shoulder instability. Furthermore, robust classification and management paradigms are required as these will serve as the effectiveness of a CDSS is dependent on quality of the knowledge base.

6. REFERENCES

BATEMAN, M., OSBORNE, S. E. & SMITH, B. E. 2019. Physiotherapy treatment for atraumatic recurrent shoulder instability: Updated results of the Derby Shoulder Instability Rehabilitation Programme. *Journal of arthroscopy and joint surgery,* 6**,** 35-41.

BRUNNER, M. C., SHEEHAN, S. E., YANKE, E. M., SITTIG, D. F., SAFDAR, N., HILL, B., LEE, K. S., ORWIN, J. F., VANNESS, D. J. & HILDEBRAND, C. J. 2020. Joint Design with Providers of Clinical Decision Support for Value-Based Advanced Shoulder Imaging. *Applied clinical informatics,* 11**,** 142-152.

CASPI, O., MCKNIGHT, P., KRUSE, L., CUNNINGHAM, V., FIGUEREDO, A. J. & SECHREST, L. 2006. Evidence-based medicine: discrepancy between perceived competence and actual performance among graduating medical students. *Medical Teacher,* 28**,** 318-325.

CHAN, M. F. & ZANG, Y. L. 2007. Nurses' perceived and actual level of diabetes mellitus knowledge: results of a cluster analysis. *Journal of Clinical Nursing,* 16**,** 234-242.

DEITCH, J., MEHLMAN, C. T., FOAD, S. L., OBBEHAT, A. & MALLORY, M. 2003. Traumatic anterior shoulder dislocation in adolescents. *The American journal of sports medicine,* 31**,** 758-763.

GERBER, C. & NYFFELER, R. W. 2002. Classification of glenohumeral joint instability. *Clin Orthop Relat Res***,** 65-76.

HCPC. 2017. *Total number of registered Physiotherapists and the number of supplementary and independent prescribers - November 2017* [Online]. Available: http://www.hcpc-uk.co.uk/resources/freedom-of-information-requests/2017/total-number-of-registered-physiotherapists-and-the-number-of-supplementary-and-independent-prescribers---november-2017/ [Accessed 25th February 2020].

JASPERS, M. W., SMEULERS, M., VERMEULEN, H. & PEUTE, L. W. 2011. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *J Am Med Inform Assoc,* 18**,** 327-34.

KUHN, J. E., HELMER, T. T., DUNN, W. R. & THROCKMORTON, V. T. 2011. Development and reliability testing of the frequency, etiology, direction, and severity (FEDS) system for classifying glenohumeral instability. *J Shoulder Elbow Surg,* 20**,** 548-56.

LANE, S., O'RAGHALLAIGH, P. & SAMMON, D. 2016. Requirements gathering: the journey. *Journal of Decision Systems,* 25**,** 302-312.

LARACCA, E., STEWART, C., POSTANS, N. & ROBERTS, A. 2014. The effects of surgical lengthening of hamstring muscles in children with cerebral palsy–the consequences of pre-operative muscle length measurement. *Gait & posture,* 39**,** 847-851.

LEROUX, T., OGILVIE-HARRIS, D., VEILLETTE, C., CHAHAL, J., DWYER, T., KHOSHBIN, A., HENRY, P., MAHOMED, N. & WASSERSTEIN, D. 2015. The epidemiology of primary anterior shoulder dislocations in patients aged 10 to 16 years. *Am J Sports Med,* 43**,** 2111-7.

LEWIS, A., KITAMURA, T. & BAYLEY, J. 2004. (ii) The classification of shoulder instability: new light through old windows! *Current Orthopaedics,* 18**,** 97-108.

LONGO, U. G., VAN DER LINDE, J. A., LOPPINI, M., COCO, V., POOLMAN, R. W. & DENARO, V. 2016. Surgical Versus Nonoperative Treatment in Patients Up to 18 Years Old With Traumatic Shoulder Instability: A Systematic Review and Quantitative Synthesis of the Literature. *Arthroscopy,* 32**,** 944-52.

MARX, R. G., MCCARTY, E. C., MONTEMURNO, T. D., ALTCHEK, D. W., CRAIG, E. V. & WARREN, R. F. 2002. Development of arthrosis following dislocation of the shoulder: A case-control study. *Journal of Shoulder and Elbow Surgery,* 11**,** 1-5.

MORODER, P., DANZINGER, V., MAZIAK, N., PLACHEL, F., PAULY, S., SCHEIBEL, M. & MINKUS, M. 2020. Characteristics of functional shoulder instability. *Journal of Shoulder and Elbow Surgery,* 29**,** 68-78.

PHILP, F., FAUX-NIGHTINGALE, A., WOOLLEY, S., DE QUINCEY, E. & PANDYAN, A. 2021. Implications for the design of a Diagnostic Decision Support System (DDSS) to reduce time and cost to diagnosis in paediatric shoulder instability. *BMC Medical Informatics and Decision Making,* 21**,** 78.

THOMAS, S. C. 1989. An approach to the repair of avulsion of the glenohumeral ligaments in the management of traumatic anterior glenohumeral instability. *The Journal of bone and joint surgery. American volume,* 71**,** 506-513.

WARBY, S. A., FORD, J. J., HAHNE, A. J., WATSON, L., BALSTER, S., LENSSEN, R. & PIZZARI, T. 2018. Comparison of 2 Exercise Rehabilitation Programs for Multidirectional Instability of the Glenohumeral Joint: A Randomized Controlled Trial. *The American journal of sports medicine,* 46**,** 87-97.

WATSON, L., WARBY, S., BALSTER, S., LENSSEN, R. & PIZZARI, T. 2016. The treatment of multidirectional instability of the shoulder with a rehabilitation program: part 1. *Shoulder & elbow,* 8**,** 271-278.

WATSON, L., WARBY, S., BALSTER, S., LENSSEN, R. & PIZZARI, T. 2017. The treatment of multidirectional instability of the shoulder with a rehabilitation programme: Part 2. *Shoulder & elbow,* 9**,** 46-53.

Analogue to Digital Telecare: Findings and Themes from a User-Centred Study to Help People Live in the Community Safely

Andrea Taylor
Glasgow School of Art
Moray, UK
a.taylor@gsa.ac.uk

Hugh Pizey
Glasgow School of Art
Glasgow, UK
h.pizey@gsa.ac.uk

Craig Whittet
Glasgow School of Art
Glasgow, UK
c.whittet@gsa.ac.uk

David Hammond
Chiptech
Lancaster, UK
david@chiptech.uk

Stephen Milne
CENSIS
Glasgow, UK
stephen.milne@censis.org.uk

Telecare services include personal alarms, home sensors and activity monitoring to enable people to remain safe and independent in their own home. Telecare has traditionally used analogue connectivity, however internationally, there is a shift to digital connectivity. This presents a rare opportunity to fundamentally redesign telecare, address current barriers to uptake, and help more people live in the community safely. This paper describes a user-centred study to design innovative digital telecare concepts, involving key stakeholders (a supplier, a manufacturer, 13 end users, 32 informal carers and 29 health and social care professionals). There are currently limited examples of digital telecare internationally. The main contributions of this paper are: an overview of key challenges and opportunities for telecare, not emphasised in existing literature within the context of the analogue to digital switchover; findings from user engagement activities, which identified issues that may be more important to users when designing telecare (e.g. self-concept) and less important (e.g. privacy); and the synthesis of ideas generated through the design process, which identified four themes that should prove useful to practitioners and researchers working in the field: community-based support, telecare you don't wear or notice, expand the use of telecare, and introduce telecare earlier.

Telecare. Technology enabled care. Assisted living technology. Telehealthcare. Older people

1. INTRODUCTION

The delivery of health and social care is changing in response to a combination of factors including: an ageing population and the associated increase in the numbers of people living longer with long-term conditions; increasing pressure on health and social health care budgets to 'do more with less'; and changes in models of health care from reactive to preventive, hospital-centred to community-based, clinician-centric to patient-centric and more recently, to consumer-centric. These changes are fuelling interest in the potential for technology, such as telecare, to support older people to remain safe and independent in their own home for longer and reduce utilisation of health services.

A range of understandings of telecare exist. This paper uses the TEC Services Association (TSA) definition: 'Telecare services include personal alarms, a wide range of home sensors (e.g. fire and flood detectors) and activity monitoring. Alerts are monitored by remote control centres that can respond quickly to emergencies' (www.tsa-voice.org.uk/support-at-home). The TSA is the industry body for telecare services in the UK. Although telecare is the term used in this paper, other terminology is used in the field such as Technology Enabled Care (TEC), assisted living technology (ALT) and telehealthcare.

Basic telecare typically comprises an alarm unit (hub); a pendant trigger worn with a neck cord, wrist strap or belt clip; and 24/7 alarm call handling. The hub incorporates an emergency alarm button and is plugged into the mains electricity supply and a home telephone line. Pressing the button on the hub alerts an Alarm Receiving Centre (ARC) that help is needed. Centre staff are able to talk with the caller through a speakerphone on the hub to decide on a course of action. For example, going straight to emergency services or involving those named on the caller's file such as a nearby informal carer (family member, friend or neighbour). Pressing the button on the pendant trigger also raises a call through the hub, provided it is within range (about 50 meters).

Enhanced telecare is more proactive than basic telecare and allows for automatic responses based on sensor information. For example, the following sensors can automatically raise an alarm call via the hub: a fall detector—an alarm is raised if the individual falls; a smoke detector; a bed sensor—an alarm is raised if an individual gets out of bed and does not return within a preset time; an epilepsy sensor—an alarm is raised when a seizure is detected; a property exit sensor—an alarm is raised when an individual's door is opened during preset times e.g. night-time. A global positioning system (GPS) is used outside the home, typically for people with dementia. An alarm is raised via satellite technology when an individual has breached a designated safe area and their whereabouts can be tracked.

Governments in most developed countries have telecare programs in place (Turner and McGee-Lennon 2013), reflecting the issue of an aging population around the world. The UK has an estimated 1.7 million end users and over 240 ARCs (TSA 2017). The main user groups for telecare are: people with telecare equipment in their home (end users); informal carers who are involved in/affected by telecare arrangements; and health and social care (H&SC) professionals e.g. home care workers and community nurses who support end users. Usually, end users of telecare are older adults. However, telecare has to be designed to support adults of all ages with varying needs and capabilities e.g. young adults with learning disabilities.

1.1. Telecare: Challenges and Opportunities

1.1.1. Analogue to Digital

In the UK, telecare is built on tried and trusted technology: telephone line connectivity delivering voice and data. However, analogue telephone services will be switched off by 2025 as the UK's telecommunications infrastructure is upgraded to digital connectivity. Analogue systems using voice-band signalling will all be affected to some degree, including telecare. Once the analogue lines have been deactivated, the current telecare infrastructure becomes less robust. The challenge is to ensure that the transition from analogue to digital is well handled, and that the lives of vulnerable people are not put at risk. However, the shift presents a rare opportunity for a fundamental redesign of telecare, rather than a 'like for like' replacement. There are few examples of digital telecare deployments in the world (FarrPoint Ltd. 2016). There are deployments based on digital technology, but these are limited in scale and number, and tend to be standalone solutions separate from the main telecare systems. A successful digital solution is therefore desirable.

1.1.2. Barriers to Uptake by End Users

A substantial body of research has shown that while many end users view telecare positively, many do not accept and/or use it as intended, and many potential users are reluctant to take it up (Hamblin 2016; Yusif, Soar, and Hafeez-Baig 2016; Stewart and McKinstry 2012; Clark and McGee-Lennon 2011; Taylor and Agamanolis 2010). Commonly cited barriers to the adoption and optimal use of telecare include: stigmatising and conspicuous equipment; unattractive equipment; the equipment is too easily activated accidentally; the cost of services; reluctance to use the equipment to disturb ('bother') call operators or informal carers; forgetting to wear equipment; a lack of knowledge or awareness about telecare; a perceived lack of need; and concerns around personal and data privacy.

Several of these problems were reported over 15 years ago (Blythe, Monk, and Doughty 2005). A contributing factor to a lack of innovation in the design of telecare is the limited interoperability between manufacturers' products: equipment supplied by one manufacturer can only 'talk' to equipment supplied by that same manufacturer. Interoperability is important because many people have a package of care (more than two items). Consequently, service providers such as local authorities are to an extent locked-in to a particular manufacturer (no competition, no innovation), and users may not have access to the equipment most suitable to their needs. Further, despite numerous studies emphasising the importance of engaging more closely with users, manufacturers have typically not done so, viewing their customer as the service providers who bulk buy and provide the equipment (Taylor et al. 2012). However, the move to digital telecare requires new technical standards, opening up new opportunities. For example, Procter, Wherton, and Greenhalgh (2018) recommend the development of more bespoke telecare solutions, as people have highly individual needs.

1.1.3. False Calls

H&SC participants in our study reported that the number of false alarm calls is as high as 80% for the pendant trigger, 87% for the fall detector, and 73% for the property exit sensor. This is important due to the cost to services to respond to calls in instances of uncertainty. A high number of false calls is also reported in much earlier research, indicating little has been achieved to address this challenge. For example, in a study with a telecare mobile warden service, the highest category of calls logged was reported as "false alarm" (Blythe, Monk, and Doughty 2005), and in a study with two ARCs in England, a teleoperator described the incoming calls from fall detectors as "all false calls" (Roberts, Mort, and Milligan 2012). Both studies reported that end users would often deliberately raise an alarm in order to have social contact. A commonly cited reason for false calls relates to technical design. For example, a study of falls in older people concluded that the sensitivity of fall detectors needs to be reduced, as they are too easily activated (Horton 2008). There is therefore a design opportunity to reduce false calls.

1.1.4. Cost Savings

Health and social care budgets are under significant pressure in most countries. Thus, telecare requires not only to be economical for end users, but also offer cost savings in health and social care budgets. In the UK, telecare can be accessed privately through product retailers or directly from suppliers or manufacturers, without recourse to local authorities, but the consumer market for it remains under-developed (Yeandle 2014). However, that some are willing to act as consumers of such products may indicate the existence of untapped demand for products and services that the telecare market could exploit more effectively than at present (Yeandle 2014). There is therefore an opportunity to stimulate consumer demand, particularly among older people aged 50+ years who hold an estimated 77% of the UK's financial worth (Centre for Ageing Better 2019).

1.2. The Innovation Challenge

This paper describes a user-centred research study conducted in the city of Glasgow, Scotland, UK. The study was funded via a mechanism that enables public sector bodies to connect with organisations from different sectors to provide innovative solutions to specific public sector challenges. In this case, the public sector body was Glasgow City Health and Social Care Partnership (GCHSCP). Health and Social Care Partnerships are partnerships between the local authority and National Health Service (NHS). The challenge was to research and design new digital telecare solutions. The study team comprised a telecare manufacture, a Higher Education Institution (HEI), and an innovation centre that specialises in Sensor and Imaging Systems and Internet of Things (IoT) technologies. This paper contributes to the body of work within HCI on assisted living technology. The main contributions are:

- Findings from user engagement activities, which identified those issues that may be more important and less important to users when designing telecare, to help ensure telecare is accepted and used.
- The synthesis of ideas generated through the design process, which identified four themes that should prove useful to other practitioners and researchers seeking to improve the acceptance and use of telecare: community-based support; telecare you don't wear or notice; expand the use of telecare; and introduce telecare earlier.
- An overview of the key challenges and opportunities for a fundamental redesign of telecare, which are not emphasised in existing literature within the context of the planned analogue to digital switchover.

2. RELATED WORK

The previous section of this paper includes related work on barriers to the uptake of telecare by end users. This section briefly describes other related research work.

2.1. AKTIVE Project

The AKTIVE (Advancing Knowledge of Telecare for Independence and Vitality in Later Life) project explored how telecare can be developed to help older adults live a full and independent life, and

benefit those caring for them, both paid and unpaid carers (AKTIVE). The AKTIVE project focused on two specific groups of older adults: those susceptible to falls and those with memory problems or dementia. The central study was called Everyday Life Analysis (ELA). ELA involved repeat research visits over six to nine months with a sample of 60 participants aged 65+ years living in two localities in England, UK. The study explored how participants experienced telecare and used or under-used equipment, including barriers to adoption.

Hamblin (2016) used the data collected from the ELA study to examine how it corresponded to an American model of 'obtrusiveness' where obtrusiveness is something which is undesirably prominent. The obtrusiveness model was created by Hensel, Demiris, and Courtney (2006) in relation to ALTs and comprises eight dimensions, each of which have subcategories (Table 1). Hamblin found that the model is largely applicable to the UK context, and identified two further issues that affect the acceptance and use of telecare: the degree of control an end user feels they have, and the information and support they receive in using their equipment. Hamblin concluded that the obtrusiveness model plus the additions (control and information) highlight important issues that can assist H&SC professionals in ensuring telecare is accepted and used.

Our study also analysed research data collected via user engagement activities using the obtrusiveness model, plus the two additions identified by Hamblin. The findings are presented in the Results and Discussion section. Our study builds on the work of Hensel, Demiris, and Courtney (2006) and Hamblin (2016) by applying the model to the Scottish context for the first time to our knowledge, and finds that the model and additions are applicable. Our study found an additional factor affecting the acceptance and use of telecare: a fear or dislike of new technology. Although a much smaller study, our study also builds on the work of Hamblin by working with a general adult population rather than specific sub-groups within the population.

Table 1: *The obtrusiveness model*

Dimension	Subcategories
Physical	*Functional dependence*
	Discomfort
	Excessive noise
	Obstruction
	Aesthetic incongruence
Usability	*Lack of user friendliness or accessibility*
	Additional demands on time and effort
Privacy	*Invasion of personal information*
	Violation of the personal space of the home
Function	*Malfunction/suboptimal performance*
	Inaccurate measurement
	Restricted distance/time away from home
	Perceived lack of usefulness
Human interaction	*Threat to replace in-person visits*
	Lack of human response in emergencies
	Negative effects on relationships
Self-concept	*Symbol of loss of independence*
	Cause of embarrassment
Routine	*Interference with daily activities*
	Acquisition of new rituals
Sustainability	*Affordability concerns*
	Concern about future needs

2.2. Athene Project

The ATHENE (Assistive Technologies for Healthy Living in Elders: Needs Assessment by Ethnography) project aimed to produce a richer understanding of the needs and lived experiences of older people, and how they and their informal carers can work with ALT stakeholders—suppliers, health and social care professionals—to co-produce ALTs (ATHENE). The research team proposed that making successful ALTs relies on collaboration, involving not only formal carers but also informal ones, whose role has gone unnoticed by technology designers. In particular, the project focused on the role of 'bricolage' (pragmatic customization, combining new with legacy devices) by informal carers, in enabling ALTs to be personalised to individual needs. The research team concluded that a new research agenda is needed, focused on solving challenges of involving users and their informal carers in the co-production of ALTs. Our study also employs a user-centred approach, involving user research and collaboration among telecare stakeholders including informal carers, to inform the design of products and services. However, our study's focus is on the challenges and opportunities arising from the transition from analogue to digital telecare.

2.3. The Development of AAL Systems

Hallewell Haslwanter, and Fitzpatrick (2017) investigated the reasons for the limited number of Ambient Assisted Living (AAL) technologies on the market. AAL systems include sensor-based systems e.g. to monitor if a person has fallen and raise an alarm, and ambient systems e.g. to detect activity in the home. The authors focused on engaging with experts involved in the development of AAL systems, rather than end users, to understand the problems development teams encounter. A total of 71 issues were identified by participants. The most important issues included: not really understanding

the needs of the user group; lack of overview of the players/which projects have been developed; and communication problems between project partners/stakeholders. Regards the first issue, the authors found the solution to be more complex than developers "just being more user centred" e.g. a lack of access to older users was identified. Our study employed a user-centred design approach to better understand the needs of users, but similarly encountered a challenge with access to end users. Specifically, a concern by GCHSCP, who facilitated access to participants, about involving vulnerable older adults in research.

3. METHOD

The study employed a user-centred design approach. Specifically, it followed a Double Diamond design process involving four stages: Discover, Define, Develop and Deliver (Design Council 2005). The study was conducted by 29 4th Year MEng Product Design Engineering students (hereafter called designers) from the HEI, supported and supervised by the study team. In addition, a researcher at the HEI with experience in telecare helped to oversee the project and synthesise the study results into research contributions. The designers were grouped into seven teams and worked on the project for 1.5 days per week for 11 weeks, involving field research and studio-based learning and teaching. The main methods of design used in the study are listed in Table 2. Ethics approval for the study was obtained from the Research Ethics Committee at the HEI and informed consent was obtained for all participants.

Table 2: *Qualitative methods of design used in the study*

Method	Description
Focus groups	*Opinions, feelings and attitudes are gauged from a group of participants about a product, service etc*
Observation	*Attentive looking and systematic recording of phenomena, including people, artefacts and environments*
Mind mapping	*Visually organizing a problem or a topic space in order to better understand it*
Affinity diagramming	*Research observations and insights are captured on post-it notes and clustered based on infinity, which form into themes*
Personas	*Personas consolidate archetypal descriptions of user behaviour patterns captured into representative profiles*
Storyboards	*Visual narratives that generate empathy and communicate the context in which a product, service etc. will be used*
Scenarios	*A narrative that explores the future use of a product or service etc. from a user's point of view*
Simulation exercises	*Deep approximations of conditions, designed to forge an immersive, empathic sense of real-life user experiences*
Prototyping	*The tangible creation of artefacts at various levels of resolution, for development and testing of ideas*

3.1. Participants

Seventy-four adults representing the main user groups for telecare described in the Introduction section participated in the study: 13 end users (11 female, 2 male); 32 informal carers (29 female, 3 male); and 29 H&SC professionals (16 female, 13 male). In addition, the designers engaged with five individuals with severe learning and physical disabilities (2 female, 3 male). These individuals did not use telecare, rather they represented extreme potential users; in considering their needs, the designers were encouraged and inspired to design solutions that are more usable by everyone. Participants were recruited via GCHSCP.

3.2. Design Process

3.2.1 Discover Phase

The Discover phase is about opening up—gathering inspiration and developing initial ideas. This phase began with a demonstration of telecare equipment by H&SC professionals with experience of prescribing and installing equipment, to support the designers to build knowledge of telecare and identify its strengths and weaknesses. Next, the designers engaged with telecare users. For logistical reasons, the designers formed into five research teams, with each team visiting one research venue. All the findings were subsequently shared among the seven design teams.

The first research team engaged with end users at a retirement housing community for older people, where homes are fitted with telecare equipment linked to the ARC in Glasgow. The next three research teams engaged with informal carers at three carer centres (one centre each), covering different areas of Glasgow. The centres provide a range of services including information, advice and respite care. The format of the sessions was focus groups within a communal space, lasting two hours. The fifth research team visited a day care service for adults with severe learning disabilities. The format of the session was semi-structured observation and discussion within a communal space. The research team observed members of staff interact with clients, clients interact with technology such as eye tracking software, and discussed clients' needs and capabilities with respect to telecare with staff.

The Discover phase also included a visit to the ARC in Glasgow to enable the designers to experience the service first hand, and to a technology demonstrator flat. The ARC visit included a demonstration of receiving an alarm call. The technology demonstrator flat is based in a multi-storey building that supports older people who are facing isolation and loneliness, or circumstances that mean they are having trouble living full and active lives. The flat was set up to showcase the different ways older people can be supported to live independently in their own homes for longer, including through use of technology.

At each of these venues, research data was collected using field notes and photographs. Data was also captured using a storyboard method at the retirement housing community and carer centres. The designers sketched visual narratives of a day in the life of individual participants, based on their accounts, to better understand their experiences. To bolster the (first-hand) research data, a set of five personas was provided by GCHSCP. The personas were built on qualitative information and portrayed users of telecare with different needs. Each persona comprised a name, a (stock) photo, a short biography and description of their behaviours and the technology they use. Staff from GCHSCP's telecare team participated in a brainstorming session, involving the generation and discussion of multiple ideas in response to each persona. The data gathered during the Discover phase was analysed using design analysis methods such as mind mapping and affinity diagramming, and structured into problem statements and initial design ideas.

3.2.2. Define Phase

The Define phase is about focusing down—synthesising a mass of ideas into a reduced number of concept designs. During this phase, the design teams defined a direction (challenge area) to focus on from all the possibilities identified in the Discover phase and identified key ideas to develop further. The process was supported by generative design methods, e.g. scenarios, which helped the teams to carefully consider how their ideas could improve people's lives, as well as build consensus and understanding among the team members. The process included desk research and site visits to the collaborating innovation centre to investigate existing technological solutions for the areas the design teams wished to address. The Define phase ended with an interim presentation attended by members of the study team. Each design team presented their research findings and concept designs for constructive feedback and a steer on which design to develop further.

3.2.3. Develop Phase

The Develop phase is about opening up—iteratively developing and testing the concept designs. The design teams revisited the retirement housing community and two of the three carer centres to gather feedback on their chosen concept designs. It was not possible to revisit the third carer centre due to scheduling difficulties. The sessions followed the format of the previous engagements (Discover Phase). The designers also engaged with telecare H&SC professionals. The format of the session was a design critique within the design studios at the HEI. Participants were divided into small groups, then on a rotating basis, each design team met with each group of participants to present their concept designs for constructive feedback (Figure 1). Participants comprised staff from the H&SCP, the carer centres, the learning disabilities service who were accompanied by three of their clients, the ARC, and the technology demonstrator flat.

Figure 1: *User engagement activity*

At each of these engagement sessions, responses to the concept designs were collected using field notes. The feedback was used to inform the iterative development of the design teams' concept designs using design methods such as prototyping and simulation exercises. For example, one of the design teams used an 'ageing body suit' located at the technology demonstrator flat to help develop their concept design. The suit restricts mobility, simulating the deteriorated agility e.g. stooped back and arthritis that is associated with aging. The Develop phase ended with a second interim presentation and critique attended by members of the study team.

3.2.4. Deliver Phase

The Deliver phase is about focusing down—finalising the resultant product, service or system. The design teams focused on final prototyping, branding (positioning and naming, logotype and colour usage), producing technical specifications and costings, and preparing presentation materials such as 3D models and information boards for a showcase event to share their research findings and final concept designs. The Deliver phase ended with the showcase, which was attended by stakeholders who had participated in the study. Following the showcase, each design team compiled a detailed report on the design process. Each report comprised approximately 100 pages (A4 size) of user research

data, visualisations (e.g. mind maps, storyboards, scenarios) and design ideas.

3.3. Synthesis of Findings and Design Ideas

As noted, a researcher with experience in telecare helped oversee the study. At the conclusion of the design process, the researcher synthesised the data collected/generated into research contributions. Specifically, the data that was synthesised comprised the findings from the user engagement activities and the ideas generated through the design process.

3.3.1. User Engagement Findings

User research data was primarily collected by the designers during the Discover and Develop phases. The data was analysed by the researcher using deductive content analysis, based on the conceptual framework (obtrusiveness model) by Hensel, Demiris, and Courtney (2006), plus the two additions (control and information) identified by Hamblin (2016). To begin, an Excel spreadsheet was created with a column for each of the codes (dimensions and subcategories) of the obtrusiveness model plus the additional codes 'control' and 'information'. The researcher then manually extracted all the user research data from the design teams' reports into the relevant codes (columns) in the Excel spreadsheet. Open coding was also used to allow for the inclusion of additional factors. This enabled the researcher to clearly identify the most important issues relating to acceptance/uptake of telecare, as reported by participants.

3.3.2. Design Ideas

Each design team delivered one final concept design i.e. there were seven concept designs in total. However, throughout the design process a multitude of ideas were generated and explored based on research. These ideas, expressed in words/phrases and images, were analysed by the researcher using affinity diagramming. To begin, the researcher re-read each of the design teams' reports several times. Each idea was then manually extracted (written) onto individual post-it notes, which were placed on a large sheet of paper. The post-it notes were then clustered based on affinity, which gave rise to the overarching themes. Finally, the theme categories were named and summarised, and checked with other members of the study team.

4. RESULTS AND DISCUSSION

4.1. User Engagement Findings

This section presents the synthesised findings from the user engagement activities. Unless reported, participants did not describe any issues relating to particular dimensions or subcategories.

4.1.1. Physical Dimension

A few participants reported issues related to excessive noise. For example, one participant commented *"I don't like the noise of the alarm"*. Several participants reported issues related to the aesthetic incongruence of telecare, in particular its unattractive and jarring appearance and *"medical aesthetic"*. One participant commented that the pull cord is *"ugly, outdated and a monstrosity"*.

4.1.2. Usability Dimension

Lack of user friendliness or accessibility was a major issue for participants. The majority of problems related to activation of devices due to physical or cognitive conditions, or learning difficulties. For example, one participant with arthritis commented *"My fingers don't work well sometimes and I can't press the button"*, and one participant with early stage dementia commented *"I forget why it's there and press it out of curiosity, then throw it into the bin when it I don't see it do anything. When the responders arrive, I get angry and confused"*. Participants also reported that devices were easily activated by accident. For example, one participant commented "*The pendant goes off all the time and it takes the alarm forever to shut off"*. The main accessibility issue related to the standard practice of installing a single hub in an end user's home. Participants reported they cannot always hear the hub and/or be understood by the ARC if they are in another room. A couple of participants reported that additional demands on time and effort were needed to charge GPS devices daily.

4.1.3. Function Dimension

Many participants reported problems with false calls from inaccurate measurement of devices, particularly from fall detectors, property exit sensors and smoke alarms. For example, one participant commented *"the smoke alarm goes off so often that I have cakes and biscuits ready for the firemen!"*. In terms of restricted distance or time away from home, many participants reported that the pendant trigger only works within range of the hub limiting its usefulness in large houses/gardens or away from home. A few participants reported a perceived lack of usefulness. For example, one participant from the retirement housing community who had deactivated their telecare system, commented *"I don't feel that I need it"*.

4.1.4. Human Interaction Dimension

A single participant reflected that asking for help from friends and family had negative effects on relationships by making her feel that she was *"no longer an equal"*. Of note, staff at the ARC in Glasgow described a number of issues related to a lack of human response in emergencies. Originally there were four call centres in Glasgow, all serving local communities where staff and residents were familiar to each other. The centres then merged into a single centre, serving the whole of Glasgow where

staff no longer have the same local knowledge of residents or geography of the area. Consequently, staff are less able to make decisions based on being knowledgeable about the client, which other research has also found to be important (Proctor et al. 2016), and travelling (response) time is longer.

4.1.5. Self-concept Dimension

Self-concept was another major issue for participants. Many participants viewed the uptake and usage of telecare as a symbol of loss of independence. For example, one participant reported that her family had organised the installation of telecare, i.e. it wasn't her choice, and that made her feel *"less independent, I'm not dead yet!"*. Participants also described telecare as a cause of embarrassment. For example, participants commented that the pendant trigger *"screams I am vulnerable"* or *"shouts I need assistance"*, and that it makes them feel like *"an inconvenience"* to ARC responders and emergency services in particular, especially in instances of false calls, as *"they have more important things to do"*.

4.1.6. Sustainability Dimension

Many participants raised affordability concerns. Historically, local authorities have provided telecare services free of charge, however many have now introduced charges. In 2013, GCHSCP introduced a weekly charge (£3) for its telecare service, which sparked a strong negative reaction: approximately 3,000 (30%) of its service users cancelled the service. Some participants commented that they couldn't afford the service e.g. *"it's a luxury I can't afford"*. Others commented that they are paying for a *"safety net"*: something that they pay for all of the time, which they might only need it some of the time in an emergency. For some, this was not worth the cost-benefit.

4.1.7. Control and Information and Support Dimensions

The control dimension was cited by a single participant as a concern, whose family had arranged for telecare to be installed in their home. This had made them feel *"inferior, it's a constant reminder of my condition"*. The information and support dimension was also an issue for some participants. For example, some participants were unclear on how to use the equipment and where it works and doesn't work. Many of the carers were unaware of the products that could help them support the person they cared for to live independently, indicating the need to improve awareness of existing products and services through health and social care channels, and as touched upon earlier, consumer channels.

Overall, the findings of the user engagement activities indicate that the model of obtrusiveness by Hensel, Demiris, and Courtney (2006) plus the additions identified by Hamblin (2016) are applicable to the Scottish context. An additional issue was identified: a fear or dislike of new technology. For example, one participant commented *"I have an iPad, but I haven't opened it and I don't have any desire to"*, and another participant with a basic (non-smart) phone commented *"I don't like touchscreens"*. Although a much smaller study, our findings are broadly similar to the study by Hamblin (2016). Neither study identified significant issues with functional dependence on telecare or obstruction, invasion of personal information, threat to replace in-person visits or interference with daily activities, suggesting these dimensions may be less important to users when designing digital telecare. Conversely, both studies identified issues with the other dimensions, in particular the self-concept dimension.

4.2. Themes

Four themes were identified from the synthesis of ideas generated through the design process:

- Community-based support
- Telecare you don't wear or notice
- Expand the use of telecare
- Introduce telecare earlier.

This section presents each theme, illustrated with one of the design teams' final concept designs, as an example of an approach that could be taken. All of the designs have an emphasis on attractive and less stigmatising products and services, and utilise low-cost technology to help ensure affordability. For example, several of the designs use LoraWAN, a wireless networking standard to support the IoT that offers long-range connectivity and low power operation cost (LoRA Alliance).

4.2.1. Community-based Support

The first theme is that of utilising community-based support to make telecare less reliant on traditional organisational requirements such as ARCs. Arguably, successful telecare relies on the existence of social networks and the availability of hands on care. Indeed, other researchers have noted a key paradox of telecare is that while it is intended to work at a distance and to be of particular value to people who do not have robust networks of co-present caring others, it will only function well when they are situated within such networks (Proctor et al. 2016). The AKTIVE project proposed that support networks for telecare users that draw on neighbours, friends or people known through local associations (e.g. a church group) give some strength to relatively "weak ties", which are theorised as especially productive of social cohesion (Yeandle 2014). Further, for policymakers and practitioners, ensuring older people have the opportunity to sustain and develop networks of weak ties may be important for future planning of support systems capable of assisting large numbers of frail older people to live safely and

independently in their communities (Yeandle 2014). A comparable conclusion of the UK Government's Whole System Demonstrator (WSD) programme—the largest randomised control trial of telecare and telehealth in the world—is that organisational requirements around telecare, such as arrangements for monitoring and responding to alarm calls, requires review if it is to become cost-effective (Newman, 2011).

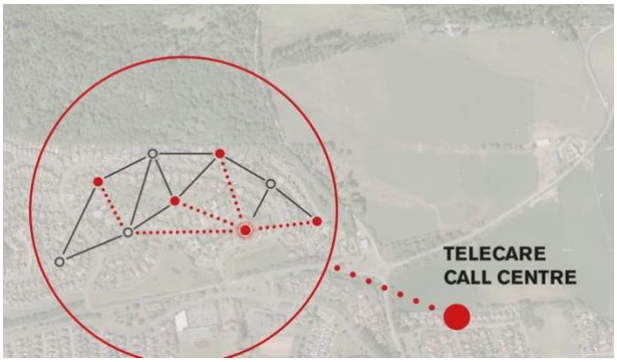


Figure 2: *A local CommuniCare network of support*

The value and potential of community-based support is illustrated in the CommuniCare concept design. CommuniCare envisions the creation of 'connected communities' where people are embedded within local networks of support for telecare. Each agreeing household has a Beacon (hub) and a set of Seeds (pendant triggers) that can be placed around the home (Figure 2). The Seeds communicate with the Beacon and each Beacon communicates with the other Beacons in the network and with the ARC via LoRaWAN and GSM SIM technologies. The more Beacons the stronger the network of support. In an emergency, the telecare user raises an alarm via the Beacon or Seed. While the community network fulfils the primary responder role, the ARC will monitor all calls and respond as necessary. The Beacon may also be used for non-emergency communications with other Beacons in the network. For example, seeking companionship or help with practical tasks. Such communications are not monitored by the ARC and could help older people combat loneliness and isolation.

4.2.2. Telecare You Don't Wear or Notice

The second theme identified was that of making telecare equipment less noticeable (obtrusive) to encourage acceptance and usage. Several of the design teams concluded that an intrinsic problem with the most common form of telecare equipment, the pendant trigger, is that end users need to be willing to wear it. However, as noted, the findings of this and other research has indicated that many end users choose not to wear their pendant and/or other wearable device (e.g. a fall detector) or they forget to put it on. In response, several of the design teams explored two directions: ambient and voice sensing technologies—removing the need for end users to be wearing a device and/or to be capable of pressing a button when help is needed; and interoperability with mainstream home technology products connected to the Internet such as smart speakers e.g. Amazon Alexa and Google Home—to help remove the noticeability of telecare equipment, and encourage people to consider purchasing consumer technologies rather than a telecare equipment package from their local authority.

Figure 3: *The Evolve vHUB (right), vSENSE (bottom), vSPEAK (top left) and vSMARTPLUG (bottom left)*

The value and potential of a non-wearable solution is illustrated in the Evolve concept design (Figure 3). Evolve envisions the creation of an unobtrusive home-based system, comprising the vHUB (hub), vSENSE, vSPEAK and vSMARTPLUG equipment. The vSENSE is a PIR sensor array that detects and monitors human movement including falls detection and property exits, as well as extreme temperature changes, and automatically raises an alarm call to the ARC e.g. in the event of a fall. vSENSE uses low resolution imaging to allay potential privacy concerns about 'spying'. The vSPEAK is an electric plug with a built-in speaker/microphone and voice recognition that allows users to voice-enable an alarm call to the ARC and communicate with Centre Staff. Finally, the vSMARTPLUG is an electric plug with a built-in speaker that connects to vSENSE, and alerts the user to plug in their GPS device for charging and/or take it with them when they exit the property.

4.2.3. Expand the Use of Telecare

Currently, telecare is aimed at people experiencing more severe declines and/or much older people, and is therefore addressing a relatively small population. The third theme was the potential to expand the use of telecare beyond those who are normally thought of as typical users. Several of the design teams explored new products and services that are desirable and functional for everyone, irrespective of age or ability. In particular, the use of telecare for overnight support (sleepovers) was explored. Sleepovers are designed to meet a range of needs including support where a person: has a significant mental health problem or learning

disability that means it is difficult for them to be alone overnight; needs a call/conversation to reassure or check-in; or might wander or leave the house. In the UK, the cost to employ a social care worker for sleepover hours has increased markedly due to a change in legislation around how staff are paid, representing a ripe opportunity area to potentially introduce telecare.

Figure 4: *The Snooze Safe bedside hub (left), and property exit sensors (right)*

The value and potential of telecare for sleepovers is illustrated in the Snooze Safe concept design (Figure 4). The system comprises a bedside hub, a property exit device, monitoring software and a mobile application (app). Snooze Safe removes the need for sleepover staff to stay overnight at the user's home. The hub uses XeTHru sensor technology to detect breathing rate and movement in order to track whether the user is in bed, and issues that are more common in people with learning disabilities such as epilepsy and sleeping problems. The hub also incorporates video calling with sleepover staff as a means to provide reassurance or check-in. For example, if the user is feeling anxious, they can make a video call and staff can offer reassurance/a human response. In the event of an emergency, the hub or property exit device automatically raises an alarm call to the ARC through the app on the user's phone using cellular and short-range wireless (Z-Wave) technologies. Monitoring data is sent to a user interface on the Snooze Safe app for sleepover staff to respond to, with the aim of keeping the user healthy in their homes for longer.

4.2.4. Introduce Telecare Earlier

The final theme relates to the earlier introduction of telecare as a pro-active choice rather than in reaction to a crisis (a 'push' situation). As described in the previous sections of this paper, there are a range of factors that affect the acceptance and use of telecare, including a fear or dislike of new technology. Some older adults are reluctant to adopt new technologies in later life that they are not familiar with, particularly when implemented in response to a time of crisis. In response, several of the design teams explored two directions: technology that end users want to acquire earlier in the life course, with a focus on prevention—so that in the event of a crisis, users are already engaged and familiar with the technology should additional equipment be required; and the concept of modularity, where the technology is flexible enough to suit the changing needs and abilities of end users as they age.

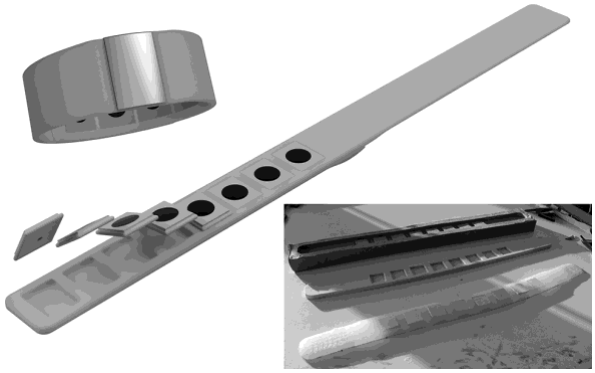


Figure 5: *The Exila band and early prototypes (right)*

The value and potential of more modular solutions is illustrated in the Exila concept design. Exila is a wearable personal trigger with eight slots for individual sensors to be inserted, a built-in speaker/microphone, and a mobile app (Figure 5). Exila combines cellular (CAT-M1 standard) and Bluetooth technologies to enable voice communication with the ARC/responder, via the wearable itself or the user's mobile phone respectively—allowing Exila to be used away from the home. Accordingly, two slots are dedicated to a SIM card and Bluetooth sensor. Examples of options for the other slots, as prototyped by the designers, include a GPS sensor for locating the end user in an emergency, heart rate sensor, temperature sensor, and accelerometer. The Exila app consolidates and visualises the data generated by the sensors, to support end users to take preventive action to maintain independence and continue living at home.

While other countries are planning digital telecare deployments, there are currently very limited examples of digital telecare internationally. Thus, there is very limited best practice and availability of digital solutions on the market (FarrPoint 2016). However, the decommissioning of analogue telephone services internationally means it is likely that a significant number of countries will need to start deploying digital telecare. This presents an important opportunity for the field of HCI to help address this immense problem and improve the acceptance/use of telecare. The four broad themes identified through our study should prove useful to other practitioners and researchers seeking to improve the acceptance and use of telecare.

4.3. Limitations

A limitation of the study is the gender imbalance in participants, where only five out of 45 (11%) participants across two of the user groups—end user, informal carers—were male. This is largely due to the very small number of men who participated in events at the retirement housing community and attended the carer centres. Therefore, the results of the study may be less representative of the target population than is desirable. However, a roughly equal number of male and female H&SC professionals participated in the study. Ideally, the study would have included more end users to increase its validity. However, as suggested by Hallewell Haslwanter and Fitzpatrick (2017), a set of personas built on user research was provided to the designers to supplement their own user research.

5. CONCLUSION

Telecare plays an important role in enabling people to remain safe and independent in their own home for longer, and most developed countries have telecare programs in place. Telecare has traditionally used analogue connectivity, however internationally, there is a shift to digital connectivity. This presents a rare opportunity to fundamentally redesign telecare, address current barriers to uptake, and help more people live in the community safely. This paper describes a user-centred study on digital telecare involving multiple stakeholders. The main contributions to the field of HCI are: an overview of the key challenges and opportunities for a fundamental redesign of telecare; findings from user engagement activities, which identified issues that may be more important and less important to users when designing telecare; and identification of four themes that should prove useful to other practitioners and researchers seeking to improve the acceptance/use of telecare: community-based support; telecare you don't wear or notice; expand the use of telecare; and introduce telecare earlier.

6. ACKNOWLEDGEMENTS

We thank all those who were involved, with particular acknowledgement to the designers of the concepts presented in this paper: Alistair Byars, Robert Mueller, Steph Parker and Samuel Watson (CommuniCare); Rebecca Jones, Isobel Leason, Africa Perez and Christopher Welch (Evolve); Jay van den Hoven, Rachael Hughson-Gill, Joe Pollard and Chiara Rossi (Exila); Jemma Brown, Scott Neill, Michael O'Donnell and Jessica Vinnels (Snooze Safe). SBRI (Small Business Research Initiative) funded the study.

7. REFERENCES

AKTIVE. "Advancing knowledge of telecare for independence and vitality in later life project". Accessed 29 April 2021. https://aktive.org.uk

ATHENE. "Assistive technologies for healthy living in elders: Needs assessment by ethnography". Accessed 29 April 2021. http://atheneproject.org.uk

Blythe, M., Monk, A., and Doughty, K. 2005. "Socially dependable design: The challenge of ageing populations for HCI." *Interacting with Computers* 17(6), 672-689. doi:10.1016/j.intcom.2005.09.005

Centre for Ageing Better. 2019. "Industrial challenge fund healthy ageing challenge framework". Accessed 29 April 2021. https://www.ageing-better.org.uk/sites/default/files/2019-02/Healthy-Ageing-Challenge-Framework.pdf

Clark, J., and McGee-Lennon, M. 2011. "A stakeholder-centred exploration of the current barriers to the uptake of home care technology in the UK." *Journal of Assistive Technologies* 5(1), 12-25. doi: 10.5042/jat.2011.0097

Design Council. 2005. "A study of the design process". Accessed 29 April 2021. https://webarchive.nationalarchives.gov.uk/20080821071137/http://www.designcouncil.org.uk/Documents/About%20design/Eleven%20Lessons/PDF%20Eleven%20Lessons_complete_study.pdf

FarrPoint Ltd. 2016. "International landscape review on the shift from analogue to digital in telecare". Accessed 29 April 2021. https://telecare.digitaloffice.scot/media/58a2fc6ed39fd083f55d4182bf88826d.pdf

Hallewell Haslwanter, J.D., and Fitzpatrick, G. 2017. Issues in the Development of AAL Systems: What experts think. *10th International Conference on PErvasive Technologies Related to Assistive Environments* (PETRA '17), 201-208. https://doi.org/10.1145/3056540.3056554

Hamblin K. 2016. "Telecare, obtrusiveness, acceptance and use: An empirical exploration." *British Journal of Occupational Therapy* 80(2), 132–138. doi:10.1177/0308022616667751

Hensel, B.K., Demiris, G., and Courtney, K.L. 2006. "Defining obtrusiveness in home telehealth technologies: A conceptual framework." *Journal of the American Medical Informatics Association* 13(4), 428-31. doi:10.1197/jamia.M2026

Horton, K. 2008. "Falls in older people: The place of telemonitoring in rehabilitation." *Journal of Rehabilitation Research and Development* 45(8), 1183-94.

LoRa Alliance. Accessed 29 April 2021. https://lora-alliance.org/about-lorawan

Martin, B., and Hannington, B. 2012. *Universal methods of design*. Rockport Publishers

Newman, S et al. 2011. “The whole System Demonstrator Programme.” Accessed 29 April 2021. https://www.city.ac.uk/__data/assets/pdf_file/0003/243066/WSD_Exec_Sum_28th_20Aug.PDF

Procter, R., Wherton, J., and Greenhalgh, T. 2018. “Hidden work and the challenges of scalability and sustainability in ambulatory assisted living.” *ACM Transactions on Computer-Human Interaction* 25(2):11, 1-26. doi:10.1145/3185591

Proctor, P., Wherton, J., Greenhalgh, T., Sugarhood, P., Rouncefield, M., and Hinder, S. 2016. “Telecare call centre work and ageing in place.” *Computer Supported Cooperative Work* 25(1), 79-105. doi:10.1007/s10606-015-9242-5

Roberts, C., Mort, M., and Milligan, C. 2012. “Calling for care: ‘Disembodied’ work, teleoperators and older people living at home.” *Sociology* 46(3), 490-506. doi:10.1177/0038038511422551

Stewart, L., and McKinstry, B. 2012. “Fear of falling and the use of telecare by older people.” *British Journal of Occupational Therapy* 75(7), 304-312. doi:10.4276/030802212X13418284515758

Taylor, A., and Agamanolis, S. 2010. “Evaluation of the user experience of a standard telecare product - the personal trigger.” *2010 Second International Conference on eHealth, Telemedicine, and Social Medicine* 51-56. doi:10.1109/eTELEMED.2010.15

Taylor, A., Bernard, L., Pizey, H., and Whittet, C. 2012. “TeleWEAR: Engaging users and suppliers of telecare in product design.” *CHI ‘12 Extended Abstracts on Human Factors in Computing Systems* 2637-2642. doi:10.1145/2212776.2223849

TSA. 2017. “A digital future for technology enabled care?” Accessed 29 April 2021. https://aace.org.uk/wp-content/uploads/2017/11/TSA-301664-Whitepaper_Oct17-120917_ONLINE-VERSION-ONLY_0.pdf

Turner, K., and McGee-Lennon, M. 2013. “Advances in telecare over the past 10 years.” *Smart Homecare Technology and TeleHealth* 1, 21–34. doi:10.2147/SHTT.S42674

Yeandle, Sue. 2014. “Frail older people and their networks of support: how does telecare fit in?” Accessed 29 April 2021. https://www.ageing.ox.ac.uk/download/61

Yusif, S., Soar, J., and Hafeez-Baig, A. 2016. “Older people, assistive technologies, and the barriers to adoption: A systematic review.” *International Journal of Medical Informatics* 94, 112-6. doi:10.1016/j.ijmedinf.2016.07.004

Usefulness Design Goals of Occupational mHealth Apps for Healthcare Workers

Nurha Yingta School of Biomedical Sciences University of West London St. Mary's Road, Ealing UK
21363511@student.uwl.ac.uk

Jose Abdelnour-Nocera University of West London St. Mary's Road, Ealing UK. ITI/Larsys, Funchal, Portugal and
Jose.Abdelnour-Nocera@uwl.ac.uk

Obed Brew School of Biomedical Sciences University of West London St. Mary's Road, Ealing UK
Obed.Brew@uwl.ac.uk

Ikram Ur Rehman School of Computing and Engineering University of West London St. Mary's Road, Ealing UK
Ikam.Rehman@uwl.ac.uk

Abstract: To improve healthcare professionals health and wellbeing at work, many available effective treatments including meditation, and workplace intervention, have been developed. However, the utilisation of these interventions is still limited. Currently, various mobile health applications (mHealth Apps) exist to assist a wide range of users with different occupational health issues, such as stress, anxiety, and burnout. Despite their advantages, post-download uptake of mHealth apps by end-users remains low. Some of the reasons for this are poor usability, irrelevant or missing user-desired features, and poor user experience. This review paper explores the usefulness of mHealth Apps to support occupational ill-health in healthcare workers. To achieve this, we developed a conceptual framework that identifies relevant usability, utility, and user experience design goals that enhance the usefulness of such mHealth apps. This paper presents a review of the literature combined with a proposed framework that identifies design goals proven to be relevant or often lacking. The review shows that occupational mHealth apps rarely fit end users' backgrounds, work contexts, and dynamics. In turn, these identified design goals will be used as assessment points with end-users in subsequent stages of our project. Expected results at the end of the project will provide an enhanced understanding of usefulness design goals that contribute to the long-term use and adoption of these apps.

Keywords*: mHealth Apps, Occupational ill-health, Usefulness, Usability, Utility, User Experience, Healthcare Workers*

1. INTRODUCTION

Occupational-related ill-health (ORIH) is a major health concern for successful economic growth. In UK, the rate of ORIH is 4.8 thousand per 100,000 workers (Hse.gov.uk., 2021), and it is commonly associated with any physical and mental health conditions that result from organisational factors as well as an imbalance of demands, skills and social support at work (Rajgopal, 2010). Current data suggests most ORIH in UK are mental health related (51% of 1.6 million cases), followed by musculoskeletal (30%) – other types of illness make up 19% (LSF, 2021).

Workers in the medical sector are at increased risk of occupational-related ill-health due to the extraordinary stressors in this environment (Ravalier, McVicar and Boichat, 2020). Stressors related to the healthcare profession include long work hours, dealing with pain, loss and emotional suffering, disease outbreak, and providing support to families (Liu *et al.*, 2020). These stressors can trigger physical and mental health issues, such as stress, burnout and anxiety.

The COVID-19 pandemic has led to feelings of unhappiness, increased stress, burnout and anxiety, particularly among healthcare workers (HCWs) (Kinman, Teoh and Harriss, 2020). Many HCWs have experienced a high prevalence of emotional burnout, physical symptoms, and work-related pressure (Barello, Palamenghi and Graffigna, 2020). This is because the fear of transmitting COVID-19 could cause HCWs to isolate from their families for months, causing the feeling of loneliness, anxiety, and depression (UK Parliament, 2020). Thus, it is crucial to design effective tools to help HCWs detect the onset of occupational-related ill-health much earlier.

The term mHealth first became popular in the early 1990s when telecommunication systems and electronic processes were used for supporting healthcare practices for the first time (Dicianno *et al.*, 2015). According to Marshall et al. (2020), an mHealth app is defined as an application program that offers health-related services through smartphones and tablets to fulfil health-specific purposes, such as stress management and prevention. Mediums for mHealth deployment are not only limited to mobile phone applications but extends to other mobile devices, such as body wearables and swallowed health monitors (Aryana, Brewster and Abdelnour-Nocera, 2018).

Despite their advantages, it remains a huge challenge to find effective mHealth apps to support occupational ill-health. Besides, not much has been done to identify and assess factors that impact on adoption, usefulness, usability, user experience and utility of mHealth apps in early detection of ORIH among healthcare workers. Usefulness refers to the degree to which users believe that using a system enhances their performance and fits with their intended tasks (Davis, 1989). Usability describes aspects that interactive products should have to allow users to continue their daily activities at work or in everyday lives without some difficulties in using the products. Utility is relevant to provide an appropriate set of functions that will allow users to carry out all their specified tasks.

Human-Computer Interaction (HCI) is concerned with understanding and designing human-centred interactive systems, where usability and user experience have been the most studied and applied concepts. Research on usefulness has been more limited (Nocera, Dunckley and Sharp, 2007; MacDonald and Atwood, 2014); therefore, reviewing factors involved with usefulness design goals is a valuable undertaking. This position paper focuses on the usefulness design goals of mobile mHealth apps to support occupational ill-health in HCWs.

The paper's outline is as follows: Section 2 provides a brief review covering related fields of HCI research, including m-Health apps in workplace setting, usefulness, usability, utility, and user experience. Section 2 also highlights a proposed framework relating to usefulness aspects based on the evidence found in the literature. This is followed in section 3 by a description of the methods involved and review findings. The paper closes by highlighting key insights to facilitate the adoption and continued use of such mHealth apps.

2. RELATED AREAS

2.1 mHealth apps in workplace settings

There are a number of studies focused on the use of mHealth app in workplace settings. For instance, de Korte *et al (*2018) have studied the use of an m-Health application called Brightr. This app was developed for workers within high-tech companies, to help them improve their health and wellbeing in the occupational setting. The results indicated that participants had muted enthusiasm about the app, as participants were unable to achieve their preferred features of an app. In another study conducted by Sarkar and colleagues (Sarkar *et al.*, 2016) it was found that participants with depression faced difficulty in using the app because they had to spend substantial among of time entering their mood data into the app. This implies that lack of proper design methods or desired utility features was presented as common reason for abandoning such mHealth apps.

The study of (Yoon *et al.*, 2021) investigated the factors that influence the adoption of m-Health apps among healthcare workers through interviews.

Results from the study showed that a few participants found the app unnecessary for promoting their mental health and wellness, as the apps offered were unnecessary for their own treatment needs. The authors suggested that the following recommendations should be considered when designing m-Health apps; (1) technical factors related to the themes perceived ease of use, convenience, security, IT support; (2) personal factors related to perceived usefulness; and lastly, the factors related to rewards and price of app.

It is thus crucial for future research to consider how mHealth apps interact with other important factors that affect the adoption and usage of such apps, for example workers' preferences, and the situations/contexts in which they work and live. The ability to leverage the needs of the end users including UX and usability in work contexts as well as desired utility features are therefore important factors enhancing the usefulness of such apps.

2.2 Defining usefulness

The usefulness of a system is a concept that has been defined and explored in different ways in disciplines such as Information System, HCI and IT (Nocera, Dunckley and Sharp, 2007). Several authors in previous studies have addressed the terms "usefulness" in different ways. For example, for Davis, usefulness is a perception of how a system enhances job performance or task completion (Davis, 1989). As Burns mentions in the context of ergonomics research, usefulness "requires that a design has the functionality required to accomplish work domain objectives" as opposed to just being easy to use (Burns *et al.*, 1997).

Most of the research on mHealth has been focused on perceived usefulness and based on acceptance models such as the Technology Acceptance Model (TAM) and e-Commerce Acceptance Model (EAM) (Schnall et al., 2016; Alsswey and Al-Samarraie, 2020). For instance, Alsswey and Al-Samarraie (2020) use the TAM model to identify the factors impacting intention of use based on culture, which feeds into the system's perceived usefulness. A similar study was found in (Schnall et al., 2016), which uses the EAM to analyse the need for mHealth to be perceived as beneficial.

Nielsen (Nielsen, 1994) defines a useful interactive system as compounded with the attribution of usability and utility. In addition to usability and utility, usefulness is influenced by the emotional feelings with a system, including enjoyment and trust to provide a richer experience of continued use (Lankton and Wilson, 2007; MacDonald and Atwood, 2014). Usefulness has also been defined as subject to interpretations grounded in the sociocultural spaces of both developers and users (Nocera, Dunckley and Sharp, 2007). Overall, the literature on usefulness reflects that this is a complex construct defined by usability, utility and user experience (UX) factors contingent on users' contexts and sociocultural backgrounds.

Designing useful mHealth apps is vital to HCWs' adoption and continued use. Considering HCWs' real-world experiences is essential to designing integrated and useful mHealth solutions (Aryana, Brewster and Abdelnour-Nocera, 2018). Therefore, to build useful mHealth apps, understanding usability, utility, and UX design goals must be considered when designing a new mobile design solution.

2.3 Proposed Framework

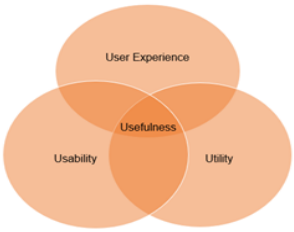


Figure 1 *A Design-Driven Conceptual Framework for Occupational mHealth*

The framework presented in Figure 1 summaries the types of goals for the usefulness of occupational mHealth apps in term of relevant usability, utility and user experience goals. These identified goals will be used as part of a questionnaire with healthcare workers in the context of a diary study of occupational mHealth app use.

3. Review Methodology and Findings

Four resources including ACM digital Library, IEEE Xplore, SAGE journal, and Google Scholar were searched in December 2020, and repeated in February 2021, to identify relevant studies. The search terms relating to mHealth apps, "usefulness", "usability", "utility", and "healthcare workers" in different Boolean permutations were used to identify relevant literature. Studies that were not journal or full conference articles, such as editorials or comments, were excluded. A total of 112 papers reporting on mHealth apps in the occupational contexts were initially identified, and after applying the exclusion criteria, 37 papers were included in the review (see Appendix for complete list of papers). The papers were then read to identify relevant usability, utility, and user experience goals. The following inclusion criteria were considered in the

screening process, as shown in Table 1. Papers that did not include all of the four criteria were excluded.

Table 1 *Inclusion criteria*

NO	Inclusion criteria
1	Studies that use mobile health application or mHealth apps
2	Studies relevant to usefulness, usability, user experience, and utility.
3	Studies that use occupational related ill-health or work-related ill-health or health
4	Studies that relevant to healthcare workers or healthcare professionals

We cannot list all reviewed papers here due to space limitations of the position paper, but we have provided at least the most important reference for each goal in Table 2.

Table 2. Identified usefulness goals of occupational mHealth apps

Usability goals	**Utility goals**	**User experience goals**
Provide contextually relevant information, which is easy to understand (Yassaee, Mettler and Winter, 2019): *Healthcare workers have said apps should reflect their own work domain context and roles.*	***Support self-help guidance and in-depth knowledge for occupational health and wellbeing (Richert, Lippke and Ziegelmann, 2011):*** *The reviewed literature reveals this as a feature that is lacking or not sufficiently developed.*	***Reinforce trust and perceived security in mHealth apps* (Byambasuren, Beller and Glasziou, 2019):** *It is necessary for users to feel confident that the system will behave as intended. This has resulted in increased collaboration with the system securely and willingly.*
***Match user expectations about the type of app: prevention or management* (Torquati *et al.*, 2018):** *Help the user access the information they need, whether the app helps them prevent or manage work-related ill-health.*	***Promote social connectedness* (Torous *et al.*, 2018):** *The app should include some communication and information sharing features such as a group collaboration among app users and with clinicians*	***Manage the performance expectancy of mHealth apps* (Odendaal *et al.*, 2015):** *The app design should be consistent with its intended goal, e.g. if it is presented as a prevention app then its features should be consistent with this aim.*
***Provide simple navigation* (Collins *et al.*, 2020)** *This goal has been found to increase adoption in workers with intense workload and constant interruptions.*	***Notify users about app updates* (e.g. app changes) (Vaghefi and Tulu, 2019)** *Inform the user when the app has some changes or updates. User should be informed about important new features*	***App should be perceived as easy to use* (Gagnon *et al.*, 2016).** *This goal has been found to increase adoption in workers with intense workload and constant interruptions.*
***Users need to recover easily from errors* (Russ and Saleem, 2018)** *This goal has been found to increase adoption in workers with intense workload and constant interruptions.*	***Gamifying app tasks wherever possible* (Bierbooms *et al.*, 2020).** *Gamification such as points and badges boosts motivation and keeps workouts fun.*	***App engagement should be rewarded (Munson and Consolvo, 2012):*** *the user needs to receive some tokens such as ribbons, coupons or cashback in return when progressing towards health goals to keep an emotional connection.*

4. Discussion and Conclusion

The findings presented here provided a review and conceptual framework that identified relevant usability, utility and user experience design goals that enhance the usefulness of mHealth apps. The identified design goals highlight not only the key dimensions of usefulness, but also the key insights needed to inform design to improve adoption and continued use of such occupational mHealth apps. For instance, in relation to usability and user experience goals, given the fast-paced nature of healthcare work, workers' everyday usage and associated experiences should be considered in the design of such apps. This is in line with previous study by (Clemmensen, Hertzum and Abdelnour-Nocera, 2020), which suggests that understanding ordinary users' experience, cognitive challenges and demands from the workplace context will ensure the design of relevant and useful mHealth solution.

Furthermore, designers or developers need to understand what content works best and how it should be implemented to increase the adoption and continued use of such apps. A future recommendation would be to implement a user-centred design in which the user will be able to express their work domain contexts and roles. This recommendation is in line with Coursaris and Kim (2011), who recommend designing an app that fits users' needs in a particular setting to improve system integrations, adoption, and loyalty.

In relation to utility goals, mHealth apps will have the potential to provide healthcare workers with a better health and wellbeing if the crucial features are effectively incorporated in such apps. This suggests that having a consideration of user desired features could lead to the increased adoption and continued used of the system (Yoon *et al.*, 2021). Future occupational mHealth apps should be more integrated with the identified usefulness design goals found to be relevant to those who work in clinical work environments or workplaces where a high level of work-related ill-health is predominant (Collins et al., 2020).

5. Conclusion

The review presented in this paper provides insightful knowledge for the design of occupational mHealth apps to enhance users' experience in order to continue using such apps. Occupational ill-health is an important issue and calls for new mHealth tools to be explored. Due to the nature of healthcare professionals work contexts and environments, future occupational mHealth apps should be designed differently following domain-relevant and distinct design goals such as those identified in this review. More importantly, Understanding HCWs experiences cognitive challenges and demands from the workplace contexts will ensure the design of relevant and useful occupational mHealth solutions. The proposed framework and goals address these aspects and are contributions to the literature on mHealth by advancing knowledge on the user-centred design of this genre of apps, focusing on healthcare workers (Yen and Bakken, 2012; Torous et al., 2018).

Expected results at the end of the project will provide an enhanced empirical understanding of usefulness design goals that contribute to the long-term use and adoption of occupational mHealth apps.

6. REFERENCES

Alsswey, A. and Al-Samarraie, H. (2020) 'Elderly users' acceptance of mHealth user interface (UI) design-based culture: the moderator role of age', *Journal on multimodal user interfaces*, 14(1), pp. 49–59. doi: 10.1007/s12193-019-00307-w.

Aryana, B., Brewster, L. and Abdelnour-Nocera, J. (2018) 'Design for mobile mental health: an exploratory review'. Available at: http://uwl.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwlVxLSwMxEB5sT16komK1ylxFWrOvNumtVBevRS97Ckl2FqS1gu3i3-8kqSie7DkhGWZI5pvHNwBZOhLDP39CQqm0jcqdTYRNnJN1qqwjQ03RMMQIuZ1SVousqnxPzf03NcYnzjdvvtw8ar9We3oVW4l81mv7UDAc6UAnE76Z6yVfBAex-uUgyh6c7JEdzqIpTuGI1mcwfgwdEsjQEN8_LD9BjNP0MRIQp2jWSKEPLpS7MVJJzkGUT6_z5yHfo6MUG-1HNP9Iqv2Scbpdai9fdgFdjuTpEjBXTghWmUwMx1VFY5yica0mQtaMA2rqw92_j706YO81HLOfD2Nc08kAutvPlm6Cqm6D2nYJdYWB.

Barello, S., Palamenghi, L. and Graffigna, G. (2020) 'Burnout and somatic symptoms among frontline healthcare professionals at the peak of the Italian COVID-19 pandemic', *Psychiatry research*, 290, p. 113129. doi: 10.1016/j.psychres.2020.113129.

Bierbooms, J. *et al.* (2020) 'Serious Games for Professional Skills: The Design of an Escape Room to Explore the Possibilities of eMental Health', in. ACM, pp. 31–35. doi: 10.1145/3393914.3395888.

Burns, C. M. *et al.* (1997) 'Towards viable, useful and usable human factors design guidance', *Applied ergonomics*, 28(5), pp. 311–322. doi: 10.1016/S0003-6870(97)00012-4.

Byambasuren, O., Beller, E. and Glasziou, P. (2019) 'Current Knowledge and Adoption of Mobile Health Apps Among Australian General Practitioners: Survey Study', *JMIR mHealth and uHealth*, 7(6), pp. e13199–e13199. doi: 10.2196/13199.

Clemmensen, T., Hertzum, M. and Abdelnour-Nocera, J. (2020) 'Ordinary User Experiences at Work: A Study of Greenhouse Growers', *ACM transactions on computer-human interaction*, 27(3), pp. 1–31. doi: 10.1145/3386089.

Collins, D. A. J. *et al.* (2020) 'A Pilot Evaluation of a Smartphone Application for Workplace Depression', *International journal of environmental research and public health*, 17(18), p. 6753. doi: 10.3390/ijerph17186753.

Davis, F. D. (1989) 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology', *MIS quarterly*, 13(3), pp. 319–340. doi: 10.2307/249008.

Dicianno, B. E. *et al.* (2015) 'Perspectives on the evolution of mobile (mHealth) technologies and application to rehabilitation', *Physical therapy*, 95(3), p. 397.

Early, J. *et al.* (2019) 'Use of Mobile Health (mHealth) Technologies and Interventions Among Community Health Workers Globally: A Scoping Review', *Health promotion practice*, 20(6), pp. 805–817. doi: 10.1177/1524839919855391.

Gagnon, M.-P. *et al.* (2016) 'm-Health adoption by healthcare professionals: a systematic review', *Journal of the American Medical Informatics Association : JAMIA*, 23(1), pp. 212–220. doi: 10.1093/jamia/ocv052.

Hse.gov.uk. (2021) *Health and safety statistics*. Available at: https://www.hse.gov.uk/statistics/> [Accessed 12 May 2021].

Kinman, G., Teoh, K. and Harriss, A. (2020) 'Supporting the well-being of healthcare workers during and after

COVID-19', *Occupational medicine (Oxford, England)*, 70. doi: 10.1093/occmed/kqaa096.
de Korte, E. M. *et al.* (2018) 'Evaluating an mHealth App for Health and Well-Being at Work: Mixed-Method Qualitative Study', *JMIR mHealth and uHealth*, 6(3), pp. e72–e72. doi: 10.2196/mhealth.6335.
Lankton, N. K. and Wilson, E. V. (2007) 'Factors Influencing Expectations of e-Health Services within a Direct-Effects Model of User Satisfaction', *E-service journal*, 5(2), pp. 85–112. doi: 10.2979/ESJ.2007.5.2.85.
Liu, N. *et al.* (2020) 'Prevalence and predictors of PTSS during COVID-19 outbreak in China hardest-hit areas: Gender differences matter', *Psychiatry research*, 287, p. 112921. doi: 10.1016/j.psychres.2020.112921.
MacDonald, C. and Atwood, M. (2014) 'What does it mean for a system to be useful?: an exploratory study of usefulness', in. ACM, pp. 885–894. doi: 10.1145/2598510.2598600.
Marshall, J. M., Dunstan, D. A. and Bartik, W. (2020) *Clinical or gimmickal: The use and effectiveness of mobile mental health apps for treating anxiety and depression*, *Australian & New Zealand Journal of Psychiatry*. (1). doi: 10.1177/0004867419876700.
Munson, S. A. and Consolvo, S. (2012) 'Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity', in *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*. IEEE, pp. 25–32. doi: 10.4108/icst.pervasivehealth.2012.248691.
Nielsen, J. (1994) 'Usability inspection methods', in *Conference companion on Human factors in computing systems*, pp. 413–414.
Nocera, J. A., Dunckley, L. and Sharp, H. (2007) 'An Approach to the Evaluation of Usefulness as a Social Construct Using Technological Frames', *International journal of human-computer interaction*, 22(1–2), pp. 153–172. doi: 10.1207/s15327590ijhc2201-02_8.
Odendaal, W. *et al.* (2015) 'Healthcare workers' perceptions and experiences on using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis', *Cochrane database of systematic reviews*, 2015(11). doi: 10.1002/14651858.CD011942.
Rajgopal, T. (2010) 'Mental well-being at the workplace', *Indian journal of occupational and environmental medicine*, 14(3), pp. 63–65. doi: 10.4103/0019-5278.75691.
Ravalier, J. M., McVicar, A. and Boichat, C. (2020) 'Work Stress in NHS Employees: A Mixed-Methods Study', *International journal of environmental research and public health*, 17(18), p. 6464. doi: 10.3390/ijerph17186464.
Richert, J., Lippke, S. and Ziegelmann, J. P. (2011) 'Intervention-engagement and its role in the effectiveness of stage-matched interventions promoting physical exercise', *Research in sports medicine (Print)*, 19(3), pp. 145–161. doi: 10.1080/15438627.2011.583164.
Sarkar, U. *et al.* (2016) 'Usability of Commercially Available Mobile Applications for Diverse Patients', *Journal of general internal medicine : JGIM*, 31(12), pp. 1417–1426. doi: 10.1007/s11606-016-3771-6.
Schnall, R. *et al.* (2016) 'A user-centered model for designing consumer mobile health (mHealth) applications (apps)', *Journal of biomedical informatics*, 60(Journal Article), pp. 243–251. doi: 10.1016/j.jbi.2016.02.002.
Torous, J. *et al.* (2018) 'Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements', *Evidence-based mental health*, 21(3), pp. 116–119.
Torquati, L. *et al.* (2018) 'Changing Diet and Physical Activity in Nurses: A Pilot Study and Process Evaluation Highlighting Challenges in Workplace Health Promotion', *Journal of Nutrition Education and Behavior*, 50(10), pp. 1015–1025. doi: 10.1016/j.jneb.2017.12.001.
Vaghefi, I. and Tulu, B. (2019) 'The Continued Use of Mobile Health Apps: Insights From a Longitudinal Study', *JMIR mHealth and uHealth*, 7(8), pp. e12983–e12983. doi: 10.2196/12983.
Yassaee, M., Mettler, T. and Winter, R. (2019) 'Principles for the design of digital occupational health systems', *Information and Organization*, 29(2), pp. 77–90. doi: https://doi.org/10.1016/j.infoandorg.2019.04.005.
Yen, P.-Y. and Bakken, S. (2012) 'Review of health information technology usability study methodologies', *Journal of the American Medical Informatics Association : JAMIA*, 19(3), pp. 413–422. doi: 10.1136/amiajnl-2010-000020.
Yoon, S. *et al.* (2021) 'Perceptions of Mobile Health Apps and Features to Support Psychosocial Well-being Among Frontline Health Care Workers Involved in the COVID-19 Pandemic Response: Qualitative Study', *Journal of medical Internet research*, 23(5), p. e26282. doi: 10.2196/26282.

APPENDIX 1- LIST OF REVIEWED PAPERS

Narváez, S. *et al.* (2016) 'Human-Centered Design of an mHealth App for the Prevention of Burnout Syndrome', *Studies in health technology and informatics*, 228, p. 215.
Hwang, W. J. and Jo, H. H. (2019) 'Evaluation of the Effectiveness of Mobile App-Based Stress-Management Program: A Randomized Controlled Trial', *International journal of environmental research and public health*, 16(21), p. 4270. doi: 10.3390/ijerph16214270. doi: 10.3390/ijerph16214270 [doi].
Huberty, J. *et al.* (2019) 'Efficacy of the Mindfulness Meditation Mobile App "Calm" to Reduce Stress Among College Students: Randomized Controlled Trial', *JMIR mHealth and uHealth*, 7(6), p. e14273. doi: 10.2196/14273 [doi].
Divall, P., Camosso-Stefinovic, J. and Baker, R. (2013) 'The use of personal digital assistants in clinical decision making by health care professionals: A systematic review', *Health informatics journal*, 19(1), pp. 16–28. doi: 10.1177/1460458212446761.
Ema, A., Nagakura, K. and Fujita, T. (2020) 'Proposal for Type Classification for Building Trust in Medical Artificial Intelligence Systems', in. ACM, pp. 251–257. doi: 10.1145/3375627.3375846.
BinDhim, N. F. *et al.* (2015) 'Depression screening via a smartphone app: cross-country user characteristics and feasibility', *Journal of the American Medical Informatics Association : JAMIA*, 22(1), pp. 29–34. doi: 10.1136/amiajnl-2014-002840.
Cochrane, K. and Schiphorst, T. (2015) 'Developing design considerations for mobile and wearable technology m-Health applications that can support recovery in mental health disorders', in. ICST, pp. 29–36. doi: 10.4108/icst.pervasivehealth.2015.259151.
Katule, N., Rivett, U. and Densmore, M. (2016) 'A Family Health App', in. ACM (ACM DEV '16), pp. 1–10. doi: 10.1145/3001913.3001920.
Mehrotra, S. *et al.* (2017) 'Unguided Mental Health Self-help Apps: Reflections on Challenges through a Clinician's Lens', *Indian journal of psychological medicine*,

39(5), pp. 707–711. doi: 10.4103/IJPSYM.IJPSYM_151_17.
Lane, N. D. *et al.* (2014) 'BeWell: Sensing Sleep, Physical Activities and Social Interactions to Promote Wellbeing', *Mobile networks and applications*. Edited by G. Fortino et al., 19(3), pp. 345–359. doi: 10.1007/s11036-013-0484-5.

Sammon, M. *et al.* (2007) 'MACCS: an industrial study of hands-free wireless communications for mobile healthcare workers', in. ACM, pp. 55–60. doi: 10.1145/1248054.1248069.
Pospos, S. *et al.* (2018) 'Web-Based Tools and Mobile Applications To Mitigate Burnout, Depression, and Suicidality Among Healthcare Students and Professionals: a Systematic Review', *Academic psychiatry*, 42(1), pp. 109–120. doi: 10.1007/s40596-017-0868-0.
Mobasheri, M. H. *et al.* (2015) 'The ownership and clinical use of smartphones by doctors and nurses in the UK: a multicentre survey study', *BMJ innovations*, 1(4), pp. 174–181. doi: 10.1136/bmjinnov-2015-000062.
Murnane, E., Huffaker, D. and Kossinets, G. (2015) 'Mobile health apps: adoption, adherence, and abandonment', in. ACM, pp. 261–264. doi: 10.1145/2800835.2800943.
Feldner, A. and Jung, Y. (2019) 'Security Issues in Mobile Healthcare Applications', in. ACM, pp. 627–630. doi: 10.1145/3366030.3366106.
Law, E. *et al.* (2009) 'Understanding, scoping and defining user experience: a survey approach', in. ACM, pp. 719–728. doi: 10.1145/1518701.1518813.
Leon, S. A. *et al.* (2007) 'Evidence-based medicine among internal medicine residents in a community hospital program using smart phones', *BMC medical informatics and decision making*, 7(1), p. 5. doi: 10.1186/1472-6947-7-5.

“I would call them, it seems faster”. The state of Telemedicine in Scotland

Georgios Sfakakis
School of Computing
Edinburgh Napier University
40404748@live.napier.ac.uk

Emilia Sobolewska
School of Computing
Edinburgh Napier University
e.sobolewska@napier.ac.uk

The novel Coronavirus pandemic has affected the way people interact with others and services. Online remote solutions have been proposed for most interactions including health care. This position paper shows that online health services are under-utilised during Covid-19, and people are often not aware of their existence. Using primary health care (General Practitioners - GPs) and the Near Me and Telecare programs in Scotland as examples, we argue that improving the usability of the GP websites can make Telemedicine more accessible and increase its adoption.

Telemedicine. Human-Computer Interaction. Covid-19. User Experience

1. INTRODUCTION

Telemedicine can play a key role during healthcare emergencies such as the Covid-19 pandemic. Using remote health services, people can receive treatment from the comfort and safety of their homes without being exposed to the danger of infection through physical contact with others. The spread of highly contagious diseases can be suppressed by reduced face-to-face interactions and thus provide health care systems with more time to prepare and respond to emergencies. For people with vulnerable immune systems or disabled people that are dependent on others, Telemedicine can seamlessly provide healthcare solutions. While researchers have highlighted the importance of Telemedicine and how it can benefit both patients and clinicians [1][2][3], so far little was done to investigate the adoption of Telemedicine during Covid-19 with an emphasis on the existing systems and people's attitude towards them. Telemedicine can be an important part of everyday healthcare and a valuable tool during a pandemic [4].
This study aimed to investigate to what extend Telemedicine is utilised during the Covid-19 outbreak in Scotland and to examine whether there are specific aspects of online health services that could be improved. The study focused on Telemedicine services offered by General Practitioners (GPs), along with existing Telemedicine programmes.

2. RELATED WORK

Health telematics is a term defined by the World Health Organization (WHO) in 1997 to describe all remote health activities or services making use of information technology to provide health education, care, disease control, and research [5]. Today, the WHO uses the term eHealth to describe all health-related activities facilitated by information and communication technologies (ICT).

The benefits of Telemedicine are manifold; it can help overcome geographical constraints to receive care in remote locations, whilst disabled people can obtain consultations and treatment without having to leave their homes and endure travelling difficulties [1]. Patients require less time for a virtual visit to the doctor compared to a physical visit, which substantially reduces the time taken off work or other daily routines. Telemedicine can be of use for the elderly, as the ageing population leads to a rise in cases of chronic diseases [2], or people with conditions such as anxiety who will experience greater amounts of emotional burdens during times such as the current pandemic [6]. It can be also applied to protect carers and patients from infected people [4].

There is an increased adoption of digital health services across Europe [7]. The study investigated 5,793 GPs from 27 European countries, and showed that most GPs use ICT systems to keep medical records electronically (93%), 78% receive laboratory reports digitally, and 66% certify sick leave remotely. The same 2019 study showed that practitioners across Europe are not utilising Telemedicine services such as remote consultations and remotely monitoring patients. Only 5% of 5,793 GPs reported they perform online consultations, and another 6% reported they have the option but choose not to. Among those who were able to offer remote health monitoring, 16% decided not to use it, which is down from 26%, in 2013.

Moreover, the study looked into the online services that GPs offer through their websites. The most frequent options were to request an appointment and repeat a prescription with 43% and 36% respectively. Meanwhile, fewer than 8% of practitioners stated that although they are able to offer these, they choose not to. While the adoption of those services was lower than 50%, there was a 13% increase for appointment request services and 11% for repeat prescription services since 2013.

Telemedicine must be well integrated into healthcare systems to be effectively used for emergencies such as the pandemic [10]. Sporadic use by a few clinicians will not enable people to adapt to Telemedicine solutions quickly. They must be given enough motivation to employ such technologies as it appears that they are not willing to accept them mostly due to the complexity and training required [11].

Despite the advantages of Digital Health and Telemedicine, some concerns have been raised. Primarily, a data protection and privacy must be addressed [12]. By digitising health services, newly created personal information about patients now has an online presence, vulnerable to potential attacks. Health organizations must make sure that any online interactions and data are secured to avoid breaches leading to potential discrimination based on a patient's health status. Laws such as the 'Genetic Information Non-discrimination Act' in the US [13] and the 'Code on Genetic Testing and Insurance' in the UK [14] protect patients from being treated differently based on their health conditions.

The importance of understanding patient's digital health literacy levels allows recommending Telemedicine options that would suit their individual skills, rather than being exclusively tied to the nature of the health condition that needs to be treated. Among the most frequent barriers between clinicians and Telemedicine services is the technical skills required, while among patients, the most frequent barriers are age, level of education, and technical literacy [15]. It is important that while making digital systems available to people, means

of improving their digital skills as well as educating themselves on health-related matters must be also provided [16]. A survey conducted in England, found that health literacy is closely connected with the use of the Internet [16]. More specifically, it appears that people with higher levels of health literacy are more likely to use the Internet for health-related purposes.

Similar to any other physical or digital service, Telemedicine solutions must offer high-quality services, and undergo a standard way for validating the quality according to its Technical, Clinical, and System aspects [17]. For instance, with the rising use of sensors in Telemedicine, it needs to ensure adequate quality and accuracy of results, to be able to correctly diagnose any conditions and avoid false health insights [18].

2.1 Telemedicine in Scotland

In 2014, the Scottish Government created the *Technology Enabled Care* (TEC) platform (tec.scot), to promote digital health services across the country. TEC aims to provide remote access to healthcare services, to improve people's health and well-being.

Along with the Scottish Government and local health communities, TEC has formed the *Digital Telecare* programme (telecare.digital.office.scot), to assist disabled people, those with health conditions, or vulnerable individuals [19]. A person subscribed for Telecare receives a technology package to monitor various conditions around the house (e.g., Fire alarms, flood and movement detectors, pill reminders, etc.). In emergencies, a designated Alarm Receiving Centre (ARC) collects calls and acts accordingly to the situation. ARC might reach out to an emergency contact of the patient, a GP, or call for an ambulance. According to the Digital Office [19], there are 128,750 people receiving telecare services across Scotland with majority (71%) being over the age of 75.

In 2018, BT announced that by 2025 it aims to replace all analogue telephone networks with digital internet protocol services (IP) [20]. The *Digital Telecare* aims to research the impact and benefits of this change to telemedicine and to establish a roadmap to gradually change all analogue telecare services (telephony connection-based) to digital services (IP-based) to be able to keep up with the changes in technology and offer high-quality telecare services.

An important part of TEC is the *"Near Me"* program (nearme.scot), allowing patients and healthcare professionals to establish a video conference session, used by NHS Scotland for remote consultations. Such a service facilitates disabled people or people vulnerable to infections to limit their physical visits to the doctor only when necessary. Also, it helps to contain the spread of infectious diseases [21]. Furthermore, it is superior to phone consultations as it adds visual stimuli where the practitioner can understand when a patient is not feeling comfortable or is having trouble understanding the conversation. The only identified drawback of Near Me is that through video conferencing sessions the level of stress and anxiety of a patient cannot be observed as well as during in-person encounters.

Telemedicine is an important developing part of the healthcare system and can be a useful tool benefiting healthcare provides as well as patients. For instance, vulnerable and disabled people could gain substantial help through Telemedicine services [22]. We aim to investigate if people are aware that such services exist, how/whether they interact with them, and to what extent the relationship with these systems was affected by the system design.

3. METHOD

This study used a mix of qualitative and quantitative methods; investigation took place online to abide by social distancing rules.

The questionnaire, to gather data regarding remote care was implemented in Google Forms (https://forms.gle/9QtQgUzAuwywZu5g8) and included 32 questions, divided into 4 parts: users' relationships with technology; awareness of Telemedicine in Scotland; the frequency and purpose of using GP's websites, and how the Coronavirus impacted that interaction. Demographic data was collected alongside the survey.

Questionnaire was shared online through social media platforms (July 2020) targeting local Scottish communities. All participants provided consent.

55 people took part in the survey, with all responses being fully completed and valid.

The content analysis was conducted to map current online resources offered by GPs, to identify services they offer, and to investigate promotion (or lack of thereof) of TEC programmes. It included 10 central GP Surgeries from the four biggest cities in Scotland, three each from Glasgow and Edinburgh, and two each from Aberdeen and Dundee.

The systems were subsequently evaluated in an online session by User Experience practitioners. Three subject experts (2 females, 1 male, aged 25-30) were asked to find basic information about the surgery (contact information, opening hours), book an appointment, and receive test results, whilst employing the think-aloud method to externalise their thoughts. They also evaluated the usability and accessibility of the systems using a SUS questionnaire [23], 10 Usability Heuristics for User Interface Design [24], as well as accessibility guidelines [25].

4. RESULTS

The questionnaire was returned by 55 participants: 22 Females, 32 Males, and one Non-binary person. The majority of the participants were between the ages of 18 and 39 (n=48), followed by seven people aged 40-69. 25 participants obtained master's degree, 17 Bachelor's, whilst the rest (n=13) completed studies at a Higher National Degree or lower. Most participants (n=32) were employed full-time, 14 were students, four were part-time employed, three self-employed and two retired or furloughed. Male participants were mostly represented in the younger age groups of 18-29 (n=16) followed by 30-39 (n=14). Female participants had a more evenly spread representation across all age groups with the largest one being the 30-39 (n=9).

Only 16 males and eight females were familiar with the term Telemedicine. Three female and five male participants stated that they are aware of the Telecare program. The 76% of the participants have not heard of Telemedicine before.

Concerning the Near Me program, only four males and six females were aware of it. The 80% of the participants are not aware of the service.

The majority of the participants (89%) stated that health care can benefit from Telemedicine. When the participants were asked if Telemedicine can benefit them, (60%) agreed or strongly agreed, 34% were undecided. The majority of the undecided were Males (n=13) with most of them (n=8) aged 18-29. Only six females stated that they were undecided.

Regarding online resources, before and during the pandemic, 22 people were instructed by their GP to engage with their website, with the majority of them (n=15) following through.

The content analysis focussed on identifying the most popular online services offered by GPs. All surgeries extended the option to repeat a prescription online, and to book an appointment, as reported by the European Commission's study [7]. Half of the websites (n=5) mentioned the option to attend a video consultation with a doctor or a nurse while three had the option to register patients online. Out of the 10, only two practices had a mention of the Telecare program.

The usability testing of the websites [24][25] concluded that most common issues include distracting pop-ups (n=4), and text-based pages (n=6) that were uneasy to search for relevant content. The expert evaluation with UX professionals shown, that though they were able to complete the tasks, they found the systems to be difficult to use, with many inconsistencies and unnecessary complexity. They have reported that pages have issues with unclear structure and navigation; they are overloaded making it difficult to browse through and find information; the lack of consistency including different colours, fonts, and font sizes used across pages, confuse the user. Furthermore, different surgeries did not share common guidelines to achieve consistency across systems.

Lastly, the experts evaluated the usability of the systems with the SUS questionnaire. The average score for the system was 34 with 68 being the minimum recommended [23].

5. RE-DESIGN AND EVALUATION

Based on the data analysis and the expert evaluation results, an alternative design was implemented, tested, and evaluated using user-centred, iterative approach.

The process included gathering requirements to define concrete scenarios, as well as prototyping and testing of initial designs. Subsequently, new design was created according to relevant findings; a high-fidelity wireframe was produced to evaluate final outcome.

The proposed design was built to facilitate an accessible system. Pale colours in darker shades of blue were used for the design to be pleasant to the eye. The blue was also picked as it is the main colour of the NHS. Black text on a white background was used to display important information while descriptive images and icons were incorporated. The font selected was Arial with 12pt – 14pt for main text and 18pt – 22pt for headlines and titles. To test the design language, a high-fidelity wireframe was created.

The prototype was tested and evaluated by 13 participants, 10 males and three females 18-39 years old, using a modified version of the Microsoft Reaction Card Method [28]. The aim was to replicate a set of frequent scenarios and tasks to test its usability and gather user feedback.

The testing session consisted of three tasks-based scenarios, including finding information about the surgery, booking an appointment, repeating a prescription, and verifying their registered requests. Prototype was fully functional.

Initially, the participants were asked to describe the system with two words (e.g., colourful, old). Most word choices suggested appropriate design, with answers such as 'professional', 'simple', classic', and 'medical' being quoted. Subsequently, participants were given a list of 20 words from the Microsoft Desirability Toolkit [28] and were prompted to pick three that best describe the system. The most popular choice was

'Understandable' and 'Consistent' two important characteristics for such a system.

Importantly, the new design improved interactions with online health services through better usability. When reviewed by experts, the proposed design achieved 88 points on the system usability scale while the average score for the existing systems that were evaluated by the experts was 34 points. This suggests that existing resources have significant space for improvements.

6. DISCUSSION AND CONCLUSIONS

Limited knowledge of existing Telemedicine services has been one of the reasons keeping its adoption at low levels [29][30], which as our study shows, is also true to Scotland. *Near Me* and *Telecare*, two major parts of Scotland's effort to offer technology-enabled health care have little or no recognition with 18.2% and 14.5% respondents respectively reporting to have heard of them. Meanwhile, more than 60% admitted that they know or might know someone who could benefit from such service. The low awareness and adoption of resources can be correlated with the frequency it is mentioned in primary care websites – only two out of the ten analysed resources referred to Telecare. Instead, the overwhelming majority of respondents reported using generic Internet content to search for health-related information (92.7%).

Despite participants having little to no knowledge of Telemedicine prior to our study, respondents have shown a positive attitude towards it, with 89% of them stating that it can benefit health care while 60% reporting that they could personally benefit from it. It appears that male participants aged 18-29 years old are the ones that admit knowing most about such systems while remaining undecided towards them. The female participants have shown a better response regarding the actual use of Telemedicine regardless of their age, which also corresponds with the numbers of Telemedicine users previously reported [8].

Based on the analysis and the experts' evaluation, most GP's websites have major usability issues. While some offer online services, weaknesses in information architecture, quality and quantity of content, navigation, consistency, and design, make systems difficult to use. A simple re-design to address those issues has shown a significand improvement in usability and information retrieval. After modifications, the SUS [23] score improved from 34 to 88/100. A common design language could be used to ensure consistency across different websites. Lastly, creating more accessible and usable resources could create better exposure to Telemedicine, and in turn increase adoption.

7. REFERENCES

[1] Kahn, J. M. (2015). Virtual Visits — Confronting the Challenges of Telemedicine. The New England Journal of Medicine, 372(18), 1684-1685.

[2] Van den Berg, Neeltje, Schumann, Maika, Kraft, Kathleen, and Hoffmann, Wolfgang. (2012). Telemedicine and telecare for older patients—A systematic review. Maturitas, 73(2), 94-114.

[3] Bashshur, R., Doarn, C. R., Frenk, J. M., Kvedar, J. C., and Woolliscroft, J. O. (2020). Telemedicine and the COVID-19 Pandemic, Lessons for the Future.

[4] Hollander, J. E., and Carr, B. G. (2020). Virtually perfect? Telemedicine for COVID-19. New England Journal of Medicine, 382(18), 1679-1681.

[5] WHO Group Consultation on Health Telematics (1997: Geneva, Switzerland). (1998). A health telematics policy in support of WHO's Health-for-all strategy for global health development: report of the WHO Group Consultation on Health Telematics, 11-16 December, Geneva, 1997. World Health Organization. https://apps.who.int/iris/handle/10665/63857 (Accessed July, 2020)

[6] Zhou, X., Snoswell, C. L., Harding, L. E., Bambling, M., Edirippulige, S., Bai, X., and Smith, A. C. (2020). The role of telehealth in reducing the mental health burden from COVID-19. Telemedicine and e-Health, 26(4), 377-379.

[7] Lupiáñez-Villanueva, F., Devaux, A., Valverde-Albacete, J., Folkvord, F., Fauli, C., Altenhofer, M., ... and Harshfield, A. (2019). Benchmarking deployment of eHealth among general practitioners.

[8] Barnett, M. L., Ray, K. N., Souza, J., and Mehrotra, A. (2018). Trends in telemedicine use in a large commercially insured population, 2005-2017. Jama, 320(20), 2147-2149.

[9] Haluza, D., Naszay, M., Stockinger, A., and Jungwirth, D. (2017). Digital natives versus digital immigrants: influence of online health information seeking on the doctor-patient relationship. Health communication, 32(11), 1342-1349.

[10] Smith, A. C., Thomas, E., Snoswell, C. L., Haydon, H., Mehrotra, A., Clemensen, J., and Caffery, L. J. (2020). Telehealth for global emergencies: Implications for coronavirus disease 2019 (COVID-19). Journal of telemedicine and telecare, 1357633X20916567.

[11] Wade, Victoria A, Eliott, Jaklin A, and Hiller, Janet E. (2014). Clinician Acceptance is the Key Factor for Sustainable Telehealth Services. Qualitative Health Research, 24(5), 682-694.

[12] Meskó, Drobni, Bényei, Gergely, and Győrffy (2017)

[13] National Human Genome Research Institute - NIH. (2018). Genetic Information Nondiscrimination Act of 2008. https://www.genome.gov/about-genomics/policy-issues/Genetic-Discrimination#gina (Accessed September, 2020)

[14] Department of Health and Social Care. (2008). Code on genetic testing and insurance. https://www.gov.uk/government/publications/code-on-genetic-testing-and-insurance (Accessed April, 2021)

[15] Scott Kruse, C., Karem, P., Shifflett, K., Vegi, L., Ravi, K., and Brooks, M. (2018). Evaluating barriers to adopting telemedicine worldwide: A systematic review. Journal of telemedicine and telecare, 24(1), 4-12.

[16] Estacio, E. V., Whittle, R., and Protheroe, J. (2019). The digital divide: examining socio-demographic factors associated with health literacy, access and use of internet to seek health information. Journal of health psychology, 24(12), 1668-1675.

[17] Mathews, S. C., McShea, M. J., Hanley, C. L., Ravitz, A., Labrique, A. B., and Cohen, A. B. (2019). Digital health: a path to validation. NPJ digital medicine, 2(1), 1-9.

[18] Shift, D. (2015, September). Technology tipping points and societal impact. In World Economic Forum Survey Report, available at: http://www3. weforum. org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015. pdf (Accessed 20.08. 2018).

[19] Digital Office Scotland. (2017). Digital Telecare. https://www.digitaloffice.scot (Accessed May, 2020)

[20] British Telecom. (2018). We are updating our home network.

[21] NHS Near Me. (15, May 2020). Using Near Me - paediatrics. [video] https://www.youtube.com/watch?v=2RxRKGxLDaU (Accessed July, 2020)

[22] Van der Vaart, R., and Drossaert, C. (2017). Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. Journal of medical Internet research, 19(1), e27.

[23] Brooke, J. (1996). SUS: a “Quick and dirty usability. Usability evaluation in industry, 189.

[24] Nielsen, J. (2014). Usability heuristics for user interface design.

[25] W3C. (2008) Web Content Accessibility Guidelines (WCAG) 2.0. https://www.w3.org/TR/2008/REC-WCAG20-20081211 (Accessed May, 2020)

[26] Friedman, Mark G, and Bryen, Diane Nelson. (2008). Web accessibility design recommendations for people with cognitive disabilities. Technology and Disability, 19(4), 205-212.

[27] Hodes, R. J., and Lindberg, D. A. (2002). Making your website senior friendly. National Institute on Aging and the National Library of Medicine.

[28] Benedek, J., and Miner, T. (2002). Measuring Desirability: New methods for evaluating desirability in a usability lab setting. Proceedings of Usability Professionals Association, 2003(8-12), 57

[29] Portnoy, J., Waller, M., and Elliott, T. (2020). Telemedicine in the Era of COVID-19. The Journal of Allergy and Clinical Immunology: In Practice, 8(5), 1489-1491.

[30] Ohannessian, R., Duong, T. A., and Odone, A. (2020). Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR public health and surveillance, 6(2), e18810.

Paper session 6: Design methods 2 Session chair: Alan Dix

Re-defining Characteristics of a Design Protagonist – Elements of Children's Design Capital

Matin Mahboob Kanafi
University of Oulu
FI-90014 Oulu, Finland
matin.mahboobkanafi@oulu.fi

Marianne Kinnula
University of Oulu
FI-90014 Oulu, Finland
marianne.kinnula@oulu.fi

Netta Iivari
University of Oulu
FI-90014 Oulu, Finland
netta.iivari@oulu.fi

The importance of child empowerment in and through design and making has been acknowledged. The notion of "child as a Design Protagonist" concerning technology has recently been introduced. We conducted a narrative literature review to examine the current understanding of what it requires from children to become a Protagonist in design. The main objective of this study is to examine the concepts associated with children's competences relevant for design, such as various capitals, skills, and capacities. We identify core concepts used as well as several gaps in this literature base. We separate the competences into 1) those that need to be nurtured in children and 2) those that children already have and bring to the design process. We propose a concept of design capital for mapping these competences of child Design Protagonists.

Design Protagonist, children, teenagers, capacity, capital, skill, asset, competency, design capital

1. INTRODUCTION

Due to the digital transformation of society, there has been an increasing concern as regards children's digital competences and their empowerment so that they can not only overcome the challenges of living in a digitalized society as technology users, but feel empowered to shape their digital futures. Children's empowerment in this sense has been emphasized in the Child-Computer Interaction (CCI) community. Researchers have extensively discussed children's participation in design process during the years (see e.g. Kawas et al. 2020; Read and Bekker 2011; Yarosh et al. 2011) and it is still a valid topic for the future CCI research (Giannakos et al. 2020). Moreover, a growing body of CCI literature examines children's empowerment in relation to design and making (Iivari and Kinnula 2018; Iversen et al. 2017; Kinnula et al. 2018; Kinnula and Iivari 2019; Schepers et al. 2018; Schepers et al. 2019), calling for 'computational empowerment' of children (Dindler et al. 2020; Iversen et al. 2018). Along these lines, the notion of 'child as a Design Protagonist' has recently been introduced (Iversen et al. 2017). It complements the discussion on children's roles in the design process (see e.g. Barendregt et al. 2016; Bekker et al. 2019; Doorn et al. 2014; Druin 2002; Kinnula et al. 2018; Landoni et al. 2016; Landoni et al. 2018; Large et al. 2003; Large et al. 2006; Schepers et al. 2018), with the Design Protagonist role particularly emphasizing the key agency of children in relation to technology and digital transformation of society.

After introduction of the Protagonist role, the literature has remained quite silent about the specific skills and competences needed in that role or how to facilitate children in becoming a Design Protagonist. There is a need to explore what the main characteristics of a Design Protagonist are and why inviting children to adopt this role is challenging (Iivari and Kinnula 2018). We scrutinize what kind of competences are associated with children's design work in the existing CCI literature. There is already plenty of valuable CCI work carried out, while a systematic mapping from this perspective is lacking. We examine the following research questions: *What kind of competences need to be developed in children regarding the Design Protagonist role*? *What kind of competences children already have that they can bring to a design process*? We answer these questions through a narrative literature review (Boell and Cecez-Kecmanovic 2015), which is a suitable literature review method for topics without established terminology. With this review, we aim first, to identify the concepts currently used to examine children's existing competences in design. This provides CCI research with useful handles with which to foster children's design and technology education and the Design Protagonist role adoption, showing children already possess valuable competences that should be better acknowledged and appreciated in design. Second, we aim to understand what competences children lack that need to be specifically nurtured to increase their agency in designing their digital futures. As a result of the analysis, we propose the concept of 'design capital' for capturing these comperences in children. With this concept, we aim at mapping developments in the field as well as moving the field forward.

The structure of this paper is as follows. First, we describe the theoretical framing of this study as a basis for the development of the design capital

concept. Second, we present the methodology employed for conducting the literature review. Third, we present the results of the literature review, including the main concepts used and their definitions. Finally, we summarize the main findings and discuss their implications for CCI research as well as their limitations and paths for future work.

2. THEORITICAL BACKGROUND

Our theoretical lens derives from Bourdieu's theory of capital, in which the notion of 'capital' refers to the resources or assets that usually take time to accumulate and are habitually distributed in different forms—economic, social, cultural, and symbolic (Bourdieu 1986). Different forms of capital are also interrelated (Bourdieu 1986), and each form of capital may potentially generate more capital (Archer et al. 2015). In this study, we focus on Bourdieu's social and cultural forms of capital. "Social capital is the aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalized relationships of mutual acquaintance or recognition—or in other words, to membership in a group" (Bourdieu 1986, p. 248). Furthermore, Bourdieu's cultural capital exists in three forms: an embodied state (e.g. language), an objectified state (e.g. books, pictures, instruments etc.), and an institutionalized state (e.g. educational qualification) (Bourdieu 1986) that have the potential to generate skills, competences, abilities and qualifications in children. In the same vein, the notion of "science capital" was coined by (Archer et al. 2014) as science-related forms of social and cultural capital to describe the different patterns of aspiration and educational participation of youth.

Within the educational field, various kinds of skills, competences, attitudes and practices have been associated with the capital concept. For example, information technology has been conceptualized as a form of cultural capital (Emmison and Frow 1998) as well as skills, competences and attitudes for information technology use (Tondeur et al. 2010), socialization into technology practice (Beckman et al. 2014) and involvement in techno-culture (Kapitzke 2000) "to explore the origins of students' technological knowledge, skills and tastes" in the school context (Beckman et al. 2018, p. 203; see also Apps et al. 2019). However, in the current literature, there is a lack of investigation akin to capital in the design context.

Our interest is generally in key competences of children in design work: on those they already have as well as on those to be developed and nurtured. When we use the term "competence" in this paper, we refer to abilities, skills, attitudes, knowledge and values: competence refers to "the ability to apply learning outcomes adequately in a defined context (...) A competence is not limited to cognitive elements (…) it also encompasses functional aspects (…) as well as interpersonal attributes and ethical values. A competence is therefore a broader concept that may actually comprise skills (as well as attitudes, knowledge etc.)" (Ananiadou and Claro 2009). The capital concept enables us to acknowledge even more variety in terms of resources and assets that children may possess and bring into the design process.

3. METHOD

We have explored the CCI literature to understand the conceptual basis of the studies on children's competences in design through a narrative literature review (Boell and Cecez-Kecmanovic 2015). Narrative literature reviews are considered useful in offering a broad perspective on the topic without established terminology (Green et al. 2017; Boell and Cecez-Kecmanovic 2015).

To collect data, the first author of this paper made a comprehensive literature search in the Scopus and the ACM databases during 9/2020-12/2020 for papers published by the end of the year 2020 in the leading venues publishing research on interaction design and children: International Journal of Child-Computer Interaction (IJCCI), International Conference on Interaction Design and Children (IDC), ACM CHI Conference on Human Factors in Computing Systems (CHI), Transactions on Computer-Human Interaction (ToCHI), Nordic Conference on Human-Computer Interaction (NordiCHI), Participatory Design Conference (PDC), Designing Interactive System (DIS) conference, Design Studies Journal, Design Journal, and Design Research Society (DRS) Conference. The review process was done in several steps to extract and select the papers. The keywords used to extract the papers were combinations of (child* OR student* OR pupil* OR kid* OR youth*) and (design* OR fabricate* OR maker* OR making* OR DYI*) and (capital* OR fund* OR asset* OR capacity* OR literacy* OR skill* OR competence* OR thinking* OR ability*) in the title, abstract, or keywords. The number of papers identified in this phase was 697.

We applied the following inclusion and exclusion criteria: 1) Non-peer reviewed papers, posters and workshops were excluded. 2) We included only studies that addressed design-related competences: they brought up the necessity of the skills for design or making activities and included actual design or making activities with children. Thus, the papers that only addressed competences during e.g., testing a design toolkit or application were excluded. 3) We excluded papers that described projects involving children with specific needs or disabilities as we assume the competences they have and need are at least partly different, even

though they are an interesting and relevant group of children to consider in the future. Built on these criteria, the final dataset included 48 papers: 23 from IDC, 11 from IJCCI, 8 from CHI, 3 from PDC, 2 from DIS and 1 from Design Journal.

We employed Mendeley software to organize the dataset. The final dataset, their abstracts and the key findings were summarized, sorted and coded regarding the competences mentioned in each included paper in a separate excel file. Two main themes were derived from the papers in a data driven manner: 1) competences that need to be developed in children, and 2) competences that children already possess and bring to the design activities. Eventually, we discussed and synthesized the main findings, with the intention to summarize the current literature, identify any possible gaps in the existing research, and provide a framework for future research (Kitchenham 2004).

4. RESULTS

4.1 Descriptive analysis

Table 1 provides an overview of the concepts used in the included papers to discuss the competences children either bring to the design process or gain or should gain through the design process.

Table 10: *Concepts used in the papers*

Concept	# of papers	Concept	# of papers
Computational thinking	*11*	*Cultural forms*	*2*
Funds of knowledge	*9*	*Historical body*	*2*
Design thinking	*8*	*Interaction order*	*2*
Maker mindset/identity	*7*	*Making literacy*	*1*
Protagonist characteristics	*6*	*Science literacy*	*1*
Digital literacy	*5*	*Science capital*	*1*
Social & cultural capital	*3*	*Funds of identity*	*1*
Tacit knowledge	*3*		

Figure 1 presents the countries in which the design activities occurred. Most of the studies were conducted in United States (n = 22) followed by Denmark (n = 5) and Finland (n = 4).

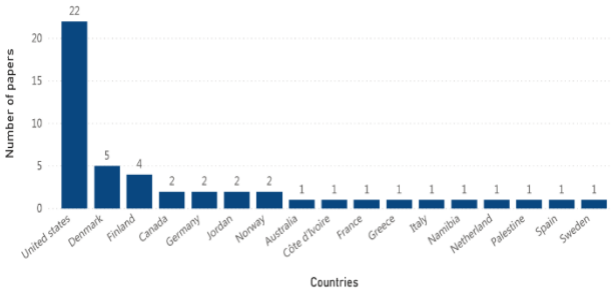


Figure 2: *Countries in the studies*

Figure 2 illustrates the age distribution of children participating in the studies. Most of the studies were conducted with 10-11-year-old children.

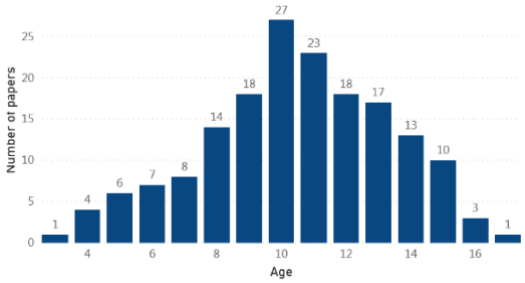


Figure 3: *Age distribution of the child participants*

Table 2 provides an overview of the publication years of the papers, showing that the number of papers has steadily grown over the years. No included paper was found before the year 2010.

Table 11: *Number of published papers in each year*

Year	#
2010	*1*
2011	*0*
2012	*1*
2013	*3*
2014	*4*
2015	*3*
2016	*6*
2017	*6*
2018	*8*
2019	*7*
2020	*9*

4.2 Narrative synthesis

Next, we present our findings structuring the text based on our research questions and the 15 central concepts used in the papers. We show how the included papers defined and utilized these concepts

and we ultimately map out the concepts to discover any possible gaps in the literature (Okoli 2015).

Table 12: *Concepts and definitions of design competences that need to be developed in children*

Concepts	Definitions
Protagonist characteristics	● *The main agent in the design process (Iversen et al. 2017)* ● *"Who steer the direction of the design process and express their individual 'voice' in it" (Södergren and Van Mechelen 2019)* ● *Independent decision making in design activity (Iversen et al. 2017)(Bonsignore et al. 2013)(Södergren and Van Mechelen 2019)* ● *Independent solving of design problem (Bonsignore et al. 2013)(Iversen et al. 2017)*
Maker Mindset / Identity	● *"A frame of thinking by which the Maker addresses problems by seeking 'do-it-yourself' solutions, a belief that they either have or can acquire the means to construct a solution, and possess a creative curiosity to seek solutions that the Maker can construct on her own" (Chu et al. 2015)* ● *Self-efficacy, motivation and interest can be seen as the core of the maker mindset (Chu et al. 2015)*
Design Thinking	● *"The ability to thoughtfully engage in design processes of digital fabrication, knowing how to act and reflect when confronted with ill-defined and complex societal problems" (Smith et al. 2015)* ● *"Design thinking is a process where a need or opportunity is identified, and a design solution is developed. The consideration of economic, environmental and social impacts that result from designed solutions are core to design thinking. Design thinking methods can be used when trying to understand a problem, generate ideas and refine a design based on evaluation and testing" (NESA 2017, p. 35)* ● *"Design thinking represents both a way of thinking and a process that can foster creative thinking. In absolute terms, design thinking stands for all the cognitive processes that a person's mind goes through when performing design" (Grammenos and Antona 2018)* ● *''A discipline that uses the designer's sensibility and methods to match people's needs with what is technologically feasible and what a viable business strategy can convert into customer value and market opportunity'' (Brown 2008)* ● *Main elements of design thinking process: "Understand, Observe, Point of View, Ideate, Prototype, Test" (Goldman et al. 2010)*
Computational thinking	● *''Thought processes involved in formulating problems so their solutions can be represented as computational steps and algorithms'' (Aho 2011)* ● *"Using abstraction and decomposition when attacking a large complex task or designing a large complex problem. It is separation on concerns. It is choosing an appropriate representation for a problem or modeling the relevant aspects of a problem to make it trackable (Wing 2006)* ● *"A problem-solving process that includes (but is not limited to) the following characteristics:* *• Formulating problems in a way that enables us to use a computer and other tools to help solve them.* *• Logically organizing and analyzing data • Representing data through abstractions such as models and simulations • Automating solutions through algorithmic thinking (a series of ordered steps) • Identifying, analyzing, and implementing possible solutions with the goal of achieving the most efficient and effective combination of steps and resources • Generalizing and transferring this problem solving process to a wide variety of problems" (ISTE 2021)*
Digital Literacy	● *"Providing children with an understanding of the use of various digital technologies, including social media, digital fabrication techniques, sensors, actuators and computing technologies" (Bekker et al. 2015)* ● *"The ability to access networked computer resources and use them" (Gilster 1997)* ● *"The ability to understand and use information in multiple formats from a wide range of sources when it is presented via computers" (Gilster 1997)* ● *" Digital literacy is associated with the ability to use computers, social media, and the Internet" (Hobbs 2010, p. 17)* ● *Three aspects of digital literacy are: 1."a set of discrete abilities or behaviors", 2. "the application of abstract mental models to activities involving digital content", 3. "engagement in a set of practices involving digital tools and media that are deeply embedded in a particular context or activity" (Meyers et al. 2013)*
Making literacy	● *"A Making-literate student functions effectively at all three levels of skills, mental models and practices…assimilate generalized ways of doing in their thinking and are able to apply them to various problems and scenarios". (Chu et al. 2017)*

4.2.1 Competences to be nurtured in children

Table 3 provides an overview of the employed concepts in the papers from the viewpoint of the competences that need to be developed in children with characteristics of the concepts or their definitions in the papers or in the original sources.

Protagonist characteristics. The concept of the Design Protagonist was coined by Iversen et al. (2017) to emphasize empowerment of children as regards technology development. The term Protagonist addresses the main agency of children when they are involved in design and problem-solving. Their main agency involves skills and critical reflection in and on design and technology development, which ultimately facilitates them to feel empowered to tackle challenges of digitized society (Iversen et al. 2017). Another recent study

provided a detailed look into the characteristics of a Design Protagonist as children were positioned into the role in a design and making project (Iivari and Kinnula 2018). Roumelioti et al. (2020) also argue that workshops in which adults act as background facilitators empower children to adopt a Protagonist role and perform as independent designers in digital making. Furthermore, Södergren and Van Mechelen (2019) introduce a design method for pre-schoolers to act as a Design Protagonist, who can make their own decision, “steer the direction of the design process and express their individual ‘voice’ in it”. Additionally, Gourlet et al. (2016) did not use the term “Protagonist” but consider independent and autonomous work of children as the main objective of their study. They mention that unpacking children’s reflective thinking during a design process can actively engage children in design and facilitate them to act autonomously. An earlier study states that inviting children to adopt a central role plays as a motivator for active participation of children (Bonsignore et al. 2013). Hence, several papers have addressed the Protagonist characteristics, but not described it comprehensively: the characteristics of a Protagonist are articulated yet quite vaguely, for instance, what kind of skills and competences children require, why some children can imagine themselves as a protagonist and others cannot, and how they can be motivated to adopt this role.

Maker Mindset/Identity. Maker mindset has been addressed as a design-related competence to be nurtured in children. According to Chu et al. (2015), one of the characteristics of making is that it is shaped through the synthesis of children’s prior knowledge, experiences and new skills. Thus, they defined a maker mindset as "a frame of thinking by which the maker address problems by seeking ‘do-it-yourself’ solutions, a belief that they either have or can acquire the means to construct a solution and possess a creative curiosity to seek solutions that the Maker can construct on her own". Furthermore, self-efficacy, motivation, and interest are considered as key elements of a maker mindset. A maker mindset can be shaped in children’s behaviour and thinking in design and making over time, which ultimately leads children to consider themselves as technology and science competent (Chu et al. 2015). In line with Chu et al. (2015) on making activities, Weibert et al. (2014, 2017) point out that making tangible artifacts improves children’s maker mindset. Malinverni et al. (2020) addressed the importance of the child becoming the “child-as-maker” to empower them through enhancing their abilities in digital fabrication and making creative use of materials. Integrating digital fabrication and making into formal learning enhances children’s maker mindset (Chu et al. 2017). In addition, as the maker mindset has been developed through Papert’s theory of Constructionism (Papert 1986), engaging children in makerspaces not only promotes their maker mindset but also eventually provides opportunities for enhancing children’s design thinking skills (Hatzigianni et al. 2020). However, nurturing a maker mindset in children is challenging as the core of the maker mindset is still vague in the literature and a deeper understanding of maker identity is needed (Iivari et al. 2018).

Design Thinking. In CCI studies, design thinking has been addressed: various attempts have been done targeting at finding useful and pertinent ways to introduce design thinking to children. Some studies focus on how design thinking can improve children’s performance, helping them in progressively understanding the complexity of design and solving problems during the design process. Smith et al. (2015) define design thinking “as the ability to thoughtfully engage in design processes of digital fabrication, knowing how to act and reflect when confronted with ill-defined and complex societal problems”. In addition, Grammenos and Antona (2018) describe that “design thinking represents both a way of thinking and a process that can foster creative thinking. In absolute terms, design thinking stands for all the cognitive processes that a person’s mind goes through when performing design.” Other studies refer to prior definitions of design thinking: for example, Hatzigianni et al. (2020) apply New South Wales Educational Standards Authority’s (NESA 2017) definition: “Design thinking is a process where a need or opportunity is identified, and a design solution is developed. The consideration of economic, environmental and social impacts that result from designed solutions are core to design thinking. Design thinking methods can be used when trying to understand a problem, generate ideas and refine a design based on evaluation and testing.” The studies maintain that design thinking enhances children’s opportunities to explore new conceptual knowledge and empowers them in their everyday digitized world. Engaging children in design thinking practices fosters creativity and innovation (Grammenos and Antona 2018; Hatzigianni et al. 2020). Furthermore, Flores (2018) mentions that when children engage in making artifacts, they can benefit from design thinking and constructionist approach. Fisher et al. (2016) and Fisher and Yefimova (2016) argue that fostering children’s design thinking facilitates children to find, generate, share, organize, and modify vital information that they need to deal with in their digitized life. Salvi (2017) argue that in poor and vulnerable locations design thinking increases awareness of children of local issues. Even if the important role of design thinking is highlighted, clear methods for introducing design thinking to children are lacking as well as the best approaches and levels of design thinking that match children’s capability and teacher’s knowledge (Bekker et al. 2015). Moreover, the possible challenges regarding

teaching design thinking to children have not yet been addressed in this literature base.

Computational Thinking (CT). Studies indicate that promoting computational thinking can empower young children in a variety of problem-solving tasks. The authors mainly utilize Wing's (2006) computational thinking definition: CT is "thinking at multiple abstractions". The studies bring up that CT is an essential skill for children to feel empowered in a digital world and actively participate in design activities. A designerly perspective to game-based design activities is also a helpful tool to promote children's CT and, more importantly, CT should be introduced based on children's funds of knowledge (Brooks and Sjöberg 2020). CT can be taught without conventional programming—rather employing children's existing knowledge, experiences and interest in a game design activity, as when children are engaged in a complex problem-solving activity, they tacitly transform their innate CT skills based on their existing interests into CT process (Lee et al. 2014). However, no method was proposed on how to teach CT built on children's funds of knowledge. Some studies argue that CT plays a key role in enhancing children's digital literacy (Tuhkala et al. 2019; Troiano et al. 2019) and that including game design practices in a constructionist learning environment lead to CT development in children (Apostolellis et al. 2014; Soleimani et al. 2016; Troiano et al. 2019; Troiano et al. 2020). Furthermore, it is reported that the utilization of spatial thinking tools in design activities nurtures children's CT skills (Soleimani et al. 2016). It is also argued that engaging children in the iterative design and debugging process which entails several constructions and deconstruction stages lead to children's CT proficiency (Litts et al. 2019). From another point of view, CT can be taught to children through tangible technologies and making artifacts (Lin and Shaer 2016) and technological tools and programming environments (Quayyum et al. 2020). Due to the importance of CT, there is an acute need to integrate it into the formal curriculum (Tuhkala et al. 2019). In addition, although CT facilitates children's problem-solving skills, it primarily encompasses computing concepts such as algorithms, decomposition and pattern recognition. CT lacks in promoting children's critical and reflective perspective on digitalization. Hence, a shift from CT to computational empowerment is argued for (Iversen et al. 2018).

Digital literacy. The concept of digital literacy has been coined by Meyers et al. (2013). Some papers studied digital literacy skills of children within the design process. According to Bekker et al. (2015) digital literacy "includes providing children with an understanding of the use of various digital technologies, including social media, digital fabrication techniques, sensors, actuators and computing technologies". Reportedly, children's digital literacy can be improved through the combination of handcrafts and digital tasks (Weibert et al. 2017), game-based learning programs (Maqsood et al. 2018), and making workshops. Eventually, children's digital literacy skills can be broadened and shaped into making literacy, which empowers children in design activity (Chu et al. 2017). Although it is reported that digital literacy can be improved through engaging children in design-based learning and digital toolkits, it can be achieved only when the digital toolkits can support learning goals and are aligned with children's various motivations for discovering modern technology (Bekker et al. 2015; Tuhkala et al. 2019).

Making Literacy. Chu et al. (2017) extended the concept of digital literacy (Meyers et al. 2013) and applied it to making activities. Thus, according to Chu et al. (2017), "A making-literate student functions effectively at all three levels of skills, mental models and practices…assimilate generalized ways of doing in their thinking and are able to apply them to various problems and scenarios". They indicate that making literacy can be taught to children if making skills and tutorials are properly adopted and children are involved in making activities.

Science Literacy. Only one paper addressed children's science literacy in design activities and no definition for science literacy was provided in this study. It is argued that utilization of a problem-solving approach as a lens in design not only promotes children's science literacy but also nurtures children's confidence and empowers them to adopt agency in a design task (Flores 2018).

4.2.2 Competences children bring to design

In Figure 4, we provide an overview of the employed concepts in the papers from the viewpoint of existing competences that children already have and bring with them to the design process. Definitions of these concepts are collected from the original sources.

Funds of Knowledge (FoK). FoK is discussed in some studies as a useful tool to enhance children's opportunities to actively participate in design. FoK is defined as "historically accumulated and culturally developed bodies of knowledge and skills essential for household or individual functioning and well-being" (Moll et al. 1992). According to Moje et al. (2004) four main themes for FoK are family, community, peer, and popular culture. It is argued that FoK do not occur in isolation and are not limited to children's individual knowledge, but rather are shaped through the network of family, community, and peers around children. These social networks provide opportunities for children to draw new resources, knowledge, and skills. Particularly, during the Covid-19 pandemic children have gained diverse knowledge distributed across their communities (Kucirkova et al. 2020). Children bring

their FoK from home and everyday experiences and integrate it into design-related science learning and

Table 13: *Concepts and definitions of competences already situated in children*

Concept	Definition
Funds of Knowledge	● *"Historically accumulated and culturally developed bodies of knowledge and skills essential for household or individual functioning and well-being" (Moll et al. 1992, p. 133)* ● *Four major themes of funds of knowledge are family, community, peer, and popular culture (Moje et al. 2004)*
Funds of Identity	● *"Historically accumulated, culturally developed, and socially distributed resources that are essential for people's self-definition, self-expression, and self-understanding (Esteban-Guitart and Moll 2014, p. 37)*
Social Capital	● *"Variety of different entities, with two elements in common: they all consist of some aspect of social structures, and they facilitate certain actions of actors-whether persons or corporate actors-within the structure" (Coleman 1988)* ● *"The aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalized relationships of mutual acquaintance and recognition—or in other words, to membership in a group—which provides each of its members with the backing of the collectively- owned capital, a "credential" which entitles them to credit, in the various senses of the word" (Bourdieu 1986, p. 248)*
Cultural capital	● *"Familiarity with the legitimate culture" in society (Bourdieu 1979)* ● *Three forms of cultural capital: embodied state, objectified state, institutionalized state (Bourdieu 1986)*
Cultural forms	● *Social constructions that are historically elaborated and connected to social activities (Saxe 1991)*
Interaction order	● *Includes people's network, social interactions, and how interaction is shaped (Goffman 1983; Scollon and Scollon 2004).*
Historical body	● *Refers to the overall aggregated experiences, histories, and knowledge of people's lives which ultimately shape and embody their behaviour (Nishida 1958; Scollon and Scollon 2004)*
Science capital	● *Science-related forms of social and cultural capital. Science capital acts as a predictive model describing the probability that a child will be drawn to a career in STEM (Archer et al. 2015).*

makerspaces. Besides the four themes mentioned above, valuable information can also be found about children's scientific FoK in informal learning such as in social media sharing apps. Children with different backgrounds mobilize their personal, cultural, and social FoK such as everyday knowledge, languages, and practices as they are engaged in sharing their everyday experiences through social media (Mills et al. 2018; Mills et al. 2019). However, a limited number of papers explicitly capitalize on different forms of FoK and apply them to investigate the value of FoK as a key resource in design activities. Some studies only mention the role of FoK in the design process without explicitly defining it (McBeath et al. 2017; Rembert et al. 2019; Yip et al. 2014). Then again, the notion of FoK provides an opportunity for design process to contemplate on transferring children's informal scientific knowledge into formal science learning contexts (Mills et al. 2018; Mills et al. 2019). In addition, it is argued that development of design activities built on children's diverse FoK not only lays the ground for children's evolving interest and motivation in design activities but also enables children to engage as designers based on their existing knowledge and competences (Brooks and Sjöberg 2020).

Furthermore, recently, a new shift has emerged in CCI that highlights a novel perspective to the notion of FoK, which is built on an asset-based approach (Kretzman and McKnight 1993). In order to empower children, the main focus in design activities should be on leveraging children's existing knowledge and competences, for instance, their parents' cultural capital, rather than their needs and lacks. (Wong-Villacres et al. 2020). The basis for this work is in 'asset-based design', a recent shift in HCI research highlighting communities' competences for sustainable development (Cho et al. 2019; Irani et al. 2018; Karusala et al. 2019; Pei and Nardi 2019; Wong-Villacres et al. 2020), the term asset being used to identify communities' existing strengths, skills and competences—rather than their shortages and needs. However, identifying these assets or varying kinds of FoK within the context of children's design capital requires work.

Tacit Knowledge. Although we found no explicit definition for tacit knowledge in the papers, some studies pointed out that children's tacit practices are associated with funds of knowledge which should be converted to explicit and formal knowledge in design activity considerations (Lee et al. 2014; Mills et al. 2018; Mills et al. 2019). Similarly, Lee et al. (2014) point out that teaching computational thinking skills to children within a design activity should be based on children's existing knowledge, experiences and interests. As children are engaged in a complex problem-solving activity, they tacitly transform their innate computational thinking skills based on their existing knowledge and interest into computational thinking process.

Funds of Identity (FoI). Only one study (Kucirkova et al. 2020) utilized the funds of identity concept and mentioned that FoI is an extension of funds of

knowledge concept. FoI is considered a useful tool to promote children's opportunities to learn within design activities. The authors refer to Esteban-Guitart and Moll's definition, which describes FoI as "historically accumulated, culturally developed, and socially distributed resources that are essential for people's self-definition, self-expression, and self-understanding" (Esteban-Guitart and Moll 2014). Children absorb resources from the network of adults around them to define themselves. That ultimately influences children to find and express themselves as an agent in design activity (Kucirkova et al. 2020).

Social and cultural capital. A limited number of the analysed papers report on social and cultural capital. In these studies, Bourdieu's notion of capital is used and linked to funds of knowledge to highlight that children's social capital is not limited to their individual knowledge, but rather distributed through networks around children. Social capital is one of the children's resources, which refers to their accumulated existing abilities and knowledge shaped through their social interactions (Wong-Villacres et al. 2020). For instance, parents as a source of social capital (Coleman 1988) are capable to facilitate learning opportunities and educational experiences for their children (Disalvo et al. 2016; Wong-Villacres et al. 2020). Furthermore, Madaio et al. (2019) utilize both Bourdieu's cultural capital and Swidler's notion of culture in action (Swidler 1986) to show how parents' values, beliefs, and aspirations for their children's education can scaffold children's knowledge in design. Moreover, children tacitly leverage their existing intangible capitals to solve their design-related problems (Wong-Villacres et al. 2020). However, children's social and cultural capital are addressed in only three papers. No explicit definitions by authors were found and only the role of parents in social and cultural capital was taken into consideration. Moreover, different forms of cultural capital such as embodied, objectified, and institutionalized (Bourdieu 1986) have so far been neglected.

Cultural forms. Children's cultural forms were reported in two studies (Horn 2013; Horn et al. 2013). Cultural forms are initially influenced by cultural funds of knowledge (Moll et al. 1992). Children's cultural forms can be seen as a key cultural resource for evoking an exploration of new competences in children within tangible design activities. The authors employ the definition of cultural forms by Saxe: cultural forms are social constructions that are historically elaborated and connected to social activities (Saxe 1991). Although cultural forms were mentioned as valuable resources, there was no discussion regarding the connection between cultural forms and related concepts.

Science Capital. Only one paper (Apostolellis et al. 2018) addresses the notion of science capital. It is argued that children's participation in informal science activities such as digital games, has an important role in generating science capital. The notion of science capital by (Archer et al. 2015) was used in the paper, describing science capital as science-related forms of social and cultural capital, which acts as a predictive model describing the probability that a child will be drawn to a career in science, engineering, technology or mathematics.

Interaction order and historical body. Nexus analysis (Scollon and Scollon 2004) was utilized in two studies (Iivari et al. 2020; Kuure et al. 2010) indicating that social action can be influenced through an intersection of children's interaction order, historical body, and discourses in place. Interaction order (Goffman 1983) resembles Bourdieu's social capital and indicates participants' networks, their social interactions, and how those are shaped. The concept of historical body (Nishida 1958) is similar to the funds of knowledge concept and refers to the overall aggregated experiences, histories, and knowledge of children, which ultimately shape and embody their behaviours. It is argued that a multidisciplinary perspective is essential in order to gain deeper insights into complex social actions of children (Iivari et al. 2020; Kuure et al. 2010).

5. DISCUSSION

5.1 Main findings

The overarching objective of this study was to identify children's competences needed for becoming a Design Protagonist. We conducted a narrative literature review and identified 48 papers that satisfied our predetermined inclusion criteria. The results indicate that various factors influence and are associated with the adoption of a Design Protagonist role. In terms of competences needed in design, several concepts were identified, including protagonist characteristics, maker mindset/identity, design thinking, computational thinking, digital literacy, computational literacy, making literacy, and science literacy. Even if a considerable number of papers mentioned the necessity of children's skills and competences, only few of them practically studied the competences needed to be nurtured within design activities with children.

An interesting insight generated in this study concerns children's competences that they bring to the design process. We identified diverse concepts including funds of knowledge, funds of identity, social and cultural capital, cultural forms, tacit knowledge, science capital, historical body, and interaction order. Even though we report a variety concepts, the limited number of papers addressing

children's existing competences indicates that these concepts have been mostly ignored in CCI research. Hence, our results indicate a number of gaps in the existing CCI literature as well as insights on how to move the field forward. These are discussed next.

5.2 Implications for CCI research

Based on our analysis, we conclude that the concepts relevant to children's competences and their relationships, differences and similarities have not yet been comprehensively or systematically examined in CCI research. We take the first step to fill this gap and extend our understanding of children's competences relevant for the role of the Design Protagonist. In the light of our findings**, we propose "design capital" as an overarching concept to capture factors relevant for the adoption of the Design Protagonist role among children**. We illustrate the aspects relevant for design capital in Figure 3: the design capital concept acts as an umbrella concept that includes a wide range of concepts found in the literature, grouping them into competences that need to be developed in children and competences that are already situated in children that they can bring to the design process. Although the competences needed among children indicate a wide variety, common for all of them is the goal of "managing and mastering digitalization". We also wish to highlight that the competences already situated in children are interrelated, as the shared component that constitutes them is a historical practice—an aspect that highlights these competences have been shaping in children through their lifetime (see Figure 3). In detail, cultural capital, funds of knowledge, funds of identity, historical body, science capital, cultural forms, and tacit knowledge refer to an individual's historically accumulated knowledge, skills, and experiences, which are usually formed through an individual's social relationships and networks of people called social capital or interaction order.

We propose design capital as a powerful new theoretical concept for structuring and guiding CCI research on children's skills and competences in design. The design capital concept might not directly help children in participating and engaging in design activities, but it assists researchers to gain a better understanding of the competences needed for a child Design Protagonist. As the CCI literature has limitations in terms of theorizing and use of theoretical concepts, this literature review contributes to CCI research by introducing a valuable new theoretical concept. The concept of capital originates from a strong theoretical basis and captures significant elements in terms of competences and assets children bring into and develop in design when acting as Design Protagonists. We maintain, in line with the existing CCI research, that children possess various kinds of existing forms of design capital that they bring to the design process, and in the design process they gain new knowledge and skills, hopefully increasing their design capital and agency to act as Design Protagonists in the future. This is in line with Bourdieu's thinking that different forms of capital are interrelated (Bourdieu 1986) and that capital may potentially generate more forms of capital (Archer et al. 2015). It is also in line with the recent introduction of the asset-based approach that emphasizes the existing resources and strengths in the support of communities (see e.g. Cho et al. 2019; Irani et al. 2018; Karusala et al. 2019; Pei and Nardi 2019; Wong-Villacres et al. 2020). However, we acknowledge that the concept of capital has not yet received a lot of attention in CCI and there is a clear need to continue this conceptual work: to clarify the different characteristics of design capital, and its dynamics, evolution and consequences in time.

We propose the design capital concept to structure and inform also CCI design practice with children. Even though the importance of children's competences such as diverse forms of capital and existing knowledge has been acknowledged within the educational context, it has mostly been ignored in design practice with children. As it seems that many CCI researchers strive to raise children to become future Design Protagonists, we need to carefully consider and develop our understandings and assumptions on the competences needed as well as existing among children. This literature review identified different facets of children's existing competences for developing design learning outcomes. Similarly, this study indicated the potential of children's existing design capital for scaffolding design developments aiming at children's empowerment. We maintain that any design activity with children should provide an opportunity for children to bring in their situated capacities and assets (Wong-Villacres et al. 2020)—the different forms of capital and funds of knowledge they possess in relation to digital technology and design (e.g. their interests, experiences, creativity and problem-solving ability). For this to happen, we recommended CCI researchers and practitioners to broadly consider the concepts identified in this study in their projects with children and try to allow and encourage children to make use of their existing competences and asserts in design as broadly as possible. Moreover, CCI researchers and practitioners should also early on in their projects reflect on what kind of competences they are particularly interested in developing in children. The concepts identified in this study have different emphasis and foci, while they also share many aspects. We particularly emphasize the need to encourage children to become Design Protagonists, who feel empowered and gain a novel understanding of digital technology and its impacts on their everyday life and society.

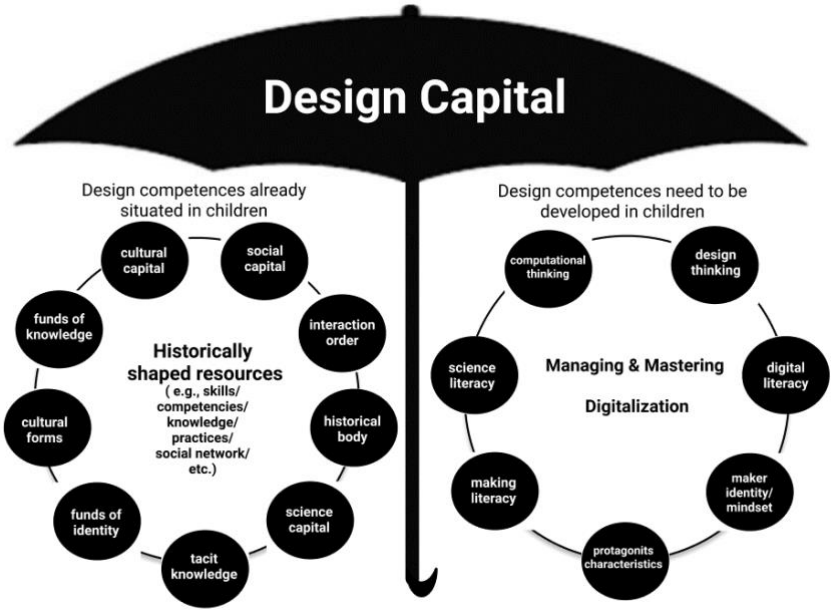


Figure 4: *Design capital*

The results of this literature review show that **further research is needed to generate more empirical evidence of children's competences in design as Design Protagonists, and of the existence, nature, and formation of children's design capital**. CCI research should also identify and examine the dynamics and interrelations of the underlying elements of design capital further (Figure 3). In addition, CCI researchers need to illuminate how these existing competences might mediate children's proficiency in design. Figure 3 lays a foundation for mapping children's competences in design and illustrates the variety of competences already existing in children as well as to be acquired However, we underscore that the competences for children to act as future Design Protagonists are not limited to the concepts mentioned in this study. Certainly, there are other potential competences that are pivotal for the Design Protagonist role for children. We acknowledge that some essential competences —such as skills, abilities, attitudes and values relating to leadership, activism and reflection, have so far been addressed in CCI research in a very limited manner.

5.3 Limitations

This literature review is limited to the search terms we used. It might be possible to find more studies on children's competences with using different terms. In addition, we excluded papers that described projects involving children with specific needs or disabilities. Additional concepts regarding children's competences might have been provided from that field of study, and that area provides interesting possibilities for future studies. We did not inquire about the effect of children's age on the concepts used. This could be considered in future studies. Several limitations are also concerned with generating our tables of concepts. Firstly, some of the children's competences were not covered in the included papers as we expected (e.g. habitus, various forms of cultural capital and design literacy). In addition, the number of papers addressing children's competences was limited and most papers that utilized the concepts did not provide comprehensive definitions for them. Some of the concepts were also employed in one or two studies only, so it prohibited us from comparing them with other studies to find similarities, differences, and relationships. A noteworthy limitation is also that this literature review only covered studies in the context of design with children, while there are interesting studies in the context of children's education as well as in the context of design with adults – both type of studies would bring additional insights on the competences associated with design. This review provided a first step addressing the developments within the core discipline of CCI. We also acknowledge that interesting insights would have been gained by reviewing the national curricula of basic education of different countries. The curricula would inform us on a variety of digital skills considered necessary for children, even if we wish to point out that many curricula remain negligent of children's design related competences still.

6. CONCLUSIONS

This paper is the first step towards a conceptual understanding of design capital that the future Design Protagonists would need. We argue that as we face the ongoing digital transformation of society and everyday life, there is a dire need for our children to grow to Design Protagonists who are empowered to critically reflect on their own technology use as well as on its use in society, and

to make and shape our digital futures. For that, they need design capital, and the CCI community is in a key role in analysing and defining what that capital means in theory and practice as well as in taking action towards nurturing it among children.

REFERENCES

Alfred V. Aho. (2011) Ubiquity Symposium: Computation and Computational Thinking. *Ubiquity*, 2011(1).

Ananiadou, K. and M. Claro (2009), 21st Century Skills and Competences for New Millennium Learners in OECD Countries. *OECD Education Working Papers,* No. 41, OECD Publishing.

Apostolellis, P., Chmiel, M. and Bowman, D.A. (2018). "Pump That Press!": Design Evaluation of Audience Inter- Action Using Collaborative Digital and Physical Games. In: *Proceedings of the 17th ACM Conference on Interaction Design and Children*. Trondheim, Norway, 19–22 June 2018. 31–42.

Apostolellis, P., Stewart, M., Frisina, C. and Kafura, D. (2014). RaBit EscApe: A Board Game for Computational Thinking. In: *Proceedings of the 2014 Conference on Interaction Design and Children*, Aarhus, Denmark, 17–20 June 2014. 349–352.

Apps, T., Beckman, K. and Bennett, S. (2019) Engaging with Bourdieu's Theory of Practice: An Empirical Tool for Exploring School Students' Technology Practice. *Technology, Pedagogy and Education*, 28 (4). 413–423.

Archer, L., Dawson, E., DeWitt, J., Seakins, A. and Wong, B. (2015) "Science Capital": A Conceptual, Methodological, and Empirical Argument for Extending Bourdieusian Notions of Capital beyond the Arts. *Journal of Research in Science Teaching*, 52 (7). 922–948.

Archer, L., Dewitt, J. and Willis, B. (2014) Adolescent Boys' Science Aspirations: Masculinity, Capital, and Power. *Journal of Research in Science Teaching*, 51 (1). 1–30.

Barendregt, W., Bekker, M.M., Börjesson, P., Eriksson, E. and Torgersson, O. (2016). The Role Definition Matrix: Creating a Shared Understanding of Children's Participation in the Design Process. In: *Proceedings of the The 15th International Conference on Interaction Design and Children*. Manchester, United Kingdom, 21-24 June 2016. 577–582.

Beckman, K., Apps, T., Bennett, S. and Lockyer, L (2018) Conceptualising Technology Practice in Education Using Bourdieu's Sociology. *Learning, Media and Technology*, Taylor & Francis, 43 (2). 197–210.

Beckman, K., Bennett, S. and Lockyer, L. (2014) Understanding Students' Use and Value of Technology for Learning. *Learning, Media and Technology*, Taylor & Francis, 39 (3). 346–367.

Bekker, T., Bakker, S., Douma, I., van der Poel, J. and Scheltenaar, K. (2015) Teaching Children Digital Literacy through Design-Based Learning with Digital Toolkits in Schools. *International Journal of Child-Computer Interaction*, 5 (Sep). 29–38.

Bekker, T., Barendregt, W., Skovbjerg, H.M., Landoni, M., Nicol, E. and Rubegni, E. (2019) Editorial Special Issue on Assumptions about the Concept of Childhood and the Roles of Children in Design. *International Journal of Child-Computer Interaction*, 19 (Mar). 89–92.

Boell, S.K. and Cecez-Kecmanovic, D. (2015) On Being "systematic" in Literature Reviews in IS. *Journal of Information Technology*, 30 (2). 161–173.

Bonsignore, E., Hansen, D., Kraus, K., Visconti, A., Ahn, J. and Druin, A. (2013). Playing for Real: Designing Alternate Reality Games for Teenagers in Learning Contexts. In: *Proceedings of the 12th International Conference on Interaction Design and Children*. New York, NY, USA, 24 - 27 June 2013. 237–246.

Bourdieu, P. (1979) *Distinction: A Social Critique of the Judgement of Taste*. Massachusetts, USA: Harvard University Press.

Bourdieu P (1986) *The forms of capital. In: Richardson JG (ed.) Handbook of Theory and Research for the Sociology of Education.* New York, Westport, CT and London: Greenwich Press, 241–258.

Brooks, E. and Sjöberg, J. (2020). A Designerly Approach as a Foundation for School Children's Computational Thinking Skills While Developing Digital Games. In: *Proceedings of the Interaction Design and Children Conference*. London, United Kingdom, 21–24 June 2020. 87–95.

Brown, T. (2008) Design Thinking. *Harvard Business Review*, 86 (6). 84.

Cho, A., Herrera, R.G., Chaidez, L. and Uriostegui, A. (2019). The "COMADRE" Project: An Asset-Based Design Approach to Connecting Low-Income Latinx Families to out-of-School Learning Opportunities. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glagsow, Scotland, UK, 4–9 May 2019. Paper 607, 1-14.

Chu, S.L., Deuermeyer, E., Martin, R., Quek, F., Berman, A., Suarez, M., Zarei, N., Nam, B. and Banigan, C. (2017). Becoming Makers: Examining 'Making' Literacy in the Elementary School Science Classroom Sharon. In:

Proceedings of the 2017 Conference on Interaction Design and Children, Stanford. CA, USA, 27–30 June 2017. 316–321.

Chu, S.L., Quek, F., Bhangaonkar, S., Ging, A.B. and Sridharamurthy, K. (2015) Making the Maker: A Means-to-an-Ends Approach to Nurturing the Maker Mindset in Elementary-Aged Children. *International Journal of Child-Computer Interaction*, 5 (Sep). 11–19.

Chu, S.L., Schlegel, R., Quek, F., Christy, A. and Chen, K. (2017). "I Make, Therefore I Am": The Effects of Curriculum-Aligned Making on Children's Self-Identity. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* Denver, CO, USA, 6–11 May 2017. 109–120.

Coleman, J.S. (1988) Social Capital in the Creation of Human Capital Source. *American Journal of Sociology*, 94. S95–S120.

Dindler, C., Smith, R. and Iversen, O.S. (2020) Computational Empowerment: Participatory Design in Education. *CoDesign*,16 (1). 66–80.

Disalvo, B., Roshan, P.K. and Morrison, B. (2016). Information Seeking Practices of Parents: Exploring Skills, Face Threats and Social Networks. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery. San Jose, California, USA, 7–12 May 2016. 623–634.

Doorn, F. van, Gielen, M. and Stappers, P.J. (2014). Children as Coresearchers. In: *Proceedings of the 2014 Conference on Interaction Design and Children*. Aarhus, Denmark, 17–20 June 2014. 237–240.

Druin, A. (2002) The Role of Children in the Design of New Technology. *Behaviour and Information Technology*, 21 (1). 1–25.

Emmison, M. and Frow, J. (1998) Information Technology as Cultural Capital. *Australian Universities' Review*, 41 (1). 41–45.

Esteban-Guitart, M. and Moll, L.C. (2014) Funds of Identity: A New Concept Based on the Funds of Knowledge Approach. *Culture and Psychology*, 20 (1). 31–48.

Fisher, K.E. and Yefimova, K. (2016). " Future ' s Butterflies :" Co-Designing ICT Wayfaring Technology with Refugee Syrian Youth. In: *Proceedings of the The 15th International Conference on Interaction Design and Children*. Manchester, United Kingdom, 21 - 24 June 2016. 25–36.

Fisher, K.E., Yefimova, K. and Bishop, A.P. (2016). Adapting Design Thinking and Cultural Probes to the Experiences of Immigrant Youth: Uncovering the Roles of Visual Media and Music in ICT Wayfaring. In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing*. San Jose, CA, USA, 07 - 12 May 2016. 859–871.

Flores, C. (2018) Problem-Based Science, a Constructionist Approach to Science Literacy in Middle School. *International Journal of Child-Computer Interaction*, 16 (June). 25–30.

Giannakos, M.N., Horn, M.S., Read, J.C. and Markopoulos, P. (2020) Movement Forward: The Continued Growth of Child–Computer Interaction Research. *International Journal of Child-Computer Interaction*, 26 (Dec).100204.

Gilster, P. (1997) *Digital Literacy*. New York, USA: Wiley Computer Pub.

Goffman, E. (1983) The Interaction Order. *American Sociological Review*, 48 (1). 1–17.

Goldman, S., Carroll, M. and Royalty, A. (2010) Destination, Imagination & the Fires within: Design Thinking in a Middle School Classroom. *International Journal of Art & Design Education*, 29 (1). 37–53.

Gourlet, P., Eveillard, L. and Dervieux, F. (2016). The Research Diary, Supporting Pupils' Reflective Thinking during Design Activities. In: *Proceedings of the The 15th International Conference on Interaction Design and Children*. Manchester, United Kingdom, 21 - 24 June 2016. 206–217.

Grammenos, D. and Antona, M. (2018) Future Designers: Introducing Creativity, Design Thinking & Design to Children. *International Journal of Child-Computer Interaction*, 16 (June).16–24.

Green, B.N., Johnson, C.D. and Adams, A. (2017) Writing Narrative Literature Reviews for Peer-Reviewed Journals: Secrets of the Trade. *Linguistic Inquiry*, 48 (2). 367–377.

Hatzigianni, M., Stevenson, M., Falloon, G., Bower, M. and Forbes, A. (2020) Young Children's Design Thinking Skills in Makerspaces. *International Journal of Child-Computer Interaction*, 27 (Mar). 100216.

Hobbs, R. (2010) *Digital and Media Literacy: A Plan of Action. A White Paper on the Digital and Media Literacy Recommendations of the Knight Commission on the Information Needs of Communities in a Democracy*. Washington, DC, USA: Aspen Institute. 1 Dupont Circle NW Suite 700.

Horn, M.S. (2013). Interaction Design, Books, and Cultural Forms. In: *Proceedings of the 12th International Conference on Interaction Design*

and Children. New York, NY, USA, 24-27 June 2013. 628–631.

Horn, M.S., AlSulaiman, S. and Koh, J. (2013). Translating Roberto to Omar: Computational Literacy, Stickerbooks, and Cultural Forms. In: *Proceedings of the 12th International Conference on Interaction Design and Children*. New York, NY, USA, 24-27 June 2013. 120–127.

Iivari, N. and Kinnula, M. (2018). Empowering Children through Design and Making: Towards Protagonist Role Adoption. In: *Proceedings of the 15th Participatory Design Conference*. Hasselt and Genk, Belgium, 20–24 August 2018. 1–12.

Iivari, N., Kinnula, M., Kuure, L. and Keisanen, T. (2020). "Arseing around Was Fun!" A- Humor as a Resource in Design and Making. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu, HI, USA, 25–30 April 2020. 1–13.

Iivari, N., Kinnula, M. and Molin-Juustila, T. (2018). You Have to Start Somewhere - Initial Meanings Making in a Design and Making Project. In: *Proceedings of the 17th ACM Conference on Interaction Design and Children*. Trondheim, Norway, 19–22 June 2018. 80–92.

Irani, A., Nelavelli, K., Hare, K., Bondal, P. and Kumar, N. (2018). Refuge Tech: An Assets-Based Approach to Refugee Resettlement. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. Montréal, QC, Canada, 21–26 April 2018. 1–6.

ISTE (2021) Operational definition of computational thinking of the international society for technology in education., available from: https://cdn.iste.org/www-root/ct-documents/computational-thinking-operational-definition-flyer.pdf?sfvrsn=2 (21 January 2021).

Iversen, O.S., Smith, R.C. and Dindler, C. (2017). Child as Protagonist: Expanding the Role of Children in Participatory Design. In: *Proceedings of the 2017 Conference on Interaction Design and Children*. Stanford, CA, USA, 27–30 June 2017. 27–37.

Iversen, O.S., Smith, R.C. and Dindler, C. (2018). From Computational Thinking to Computational Empowerment: A 21st Century PD Agenda. In: *Proceedings of the 15th Participatory Design Conference: Full Papers*, Hasselt and Genk, Belgium, 20–24 August 2018. 1–11.

Kapitzke, C. (2000) Information Technology as Cultural Capital: Shifting the Boundaries of Power. *Education and Information Technologies*, 5 (1). 49–62.

Karusala, N., Bhalla, A. and Kumar, N. (2019). Privacy, Patriarchy, and Participation on Social Media. In: *Proceedings of the 2019 on Designing Interactive Systems Conference*, San Diego, CA, USA, 23–28 June 2019. 511–526.

Kawas, S., Yuan, Y., Dewitt, A., Jin, Q., Kirchner, S., Bilger, A., Grantham, E., Kientz, J.A., Tartaro, A. and Yarosh, S. (2020). Another Decade of IDC Research: Examining and Reflecting on Values and Ethics. In: *Proceedings of the Interaction Design and Children Conference*. London, United Kingdom, 21–24 june 2020. 205–215.

Kinnula, M. and Iivari, N. (2019). Empowered to Make a Change: Guidelines for Empowering the Young Generation in and through Digital Technology Design. In: *Proceedings of the FabLearn Europe 2019 Conference.* New York, NY, USA, 28-29 May 2019. 1–8.

Kinnula, M., Iivari, N., Isomursu, M. and Kinnula, H. (2018) Socializers, Achievers or Both? Value-Based Roles of Children in Technology Design Projects. *International Journal of Child-Computer Interaction*, 12 (Sep). 39–49.

Kitchenham, B. (2004) Procedures for Performing Systematic Reviews. *Keele, UK, Keele University*, 33 (2004). 1–26.

Kretzman, J. P., and McKnight, J. (1993) Introduction to Building Communities from the Inside Out: A Path Toward Finding and Mobilizing a Community's Assets. *Institute for Policy Research, Northwestern University,* Evanston, IL, USA.

Kucirkova, N., Evertsen-Stanghelle, C., Studsrød, I., Jensen, I.B. and Størksen, I. (2020) Lessons for Child–Computer Interaction Studies Following the Research Challenges during the Covid-19 Pandemic. *International Journal of Child-Computer Interaction*, 26 (Dec). 100203.

Kuure, L., Halkola, E., Iivari, N., Kinnula, M. and Molin-Juustila, T. (2010). Children Imitate! Appreciating Recycling in Participatory Design with Children. In: *Proceedings of the 11th Biennial Participatory Design Conference*. Sydney, Australia, 29 Nov 2010. 131–140.

Landoni, M., Rubegni, E. and Nicol, E. (2018) A Comparative Study into How Pupils Can Play Different Roles in Co-Design Activities. *International Journal of Child-Computer Interaction*, 17 (Sep). 28–38.

Landoni, M., Rubegni, E., Nicol, E. and Read, J. (2016). How Many Roles Can Children Play?. In: *Proceedings of the The 15th International Conference on Interaction Design and Children*. Manchester, United Kingdom, 21-24 June 2016. 720–725.

Large, A., Beheshti, J., Nesset, V. and Bowler, L. (2003). Children as Designers of Web Portals. In: *Proceedings of the American Society for*

Information Science and Technology. 40 (1), 142–149.

Large, A., Nesset, V., Beheshti, J. and Bowler, L. (2006) "Bonded Design": A Novel Approach to Intergenerational Information Technology Design. *Library and Information Science Research*, 28 (1). 64–82.

Lee, T.Y., Mauriello, M.L., Ahn, J. and Bederson, B.B. (2014) CTArcade: Computational Thinking with Games in School Age Children. *International Journal of Child-Computer Interaction*, 2 (1). 26–33.

Lin, V. and Shaer, O. (2016). Beyond the Lab: Using Technology Toys to Engage South African Youth in Computational Thinking. In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. San Jose, CA, USA, 07-12 May 2016. 655–661.

Litts, B.K., Chauhan, A., Mortensen, C.K. and Matthias, K. (2019). "I'm Drowning in Squirrels!": How Children Embody and Debug Computational Algorithms through Designing Mixed Reality Games. In: *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. Boise, Idaho, USA, 12-15 June 2019. 267–273.

Madaio, M.A., Tanoh, F., Seri, A.B., Jasinska, K. and Ogan, A. (2019). "Everyone Brings Their Grain of Salt": Designing for Low-Literate Parental Engagement with Children's Literacy in Côte d'Ivoire. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow, Scotland Uk, 4–9 May 2019. 1–14.

Malinverni, L., Schaper, M.M. and Valero, C. (2020) Relating to Materials in Digital Fabrication: Transform Materials to Transform Yourself. *International Journal of Child-Computer Interaction*, 23–24 (June). 100166.

Maqsood, S., Mekhail, C. and Chiasson, S. (2018). A Day in the Life of Jos: A Web-Based Game to Increase Children's Digital Literacy. In: *Proceedings of the 17th ACM Conference on Interaction Design and Children*. Trondheim, Norway, 19–22 June 2018. 241–252.

McBeath, J.K., Durán, R.P. and Harlow, D.B. (2017). Not My Gumdrop Buttons! Youth Tool Use in Designing an Electronic Shrek-Themed Bean Bag Toss. In: *Proceedings of the 2017 Conference on Interaction Design and Children*. Stanford, CA, USA, 27–30 June 2017. 61–72.

Meyers, E.M., Erickson, I. and Small, R. V. (2013) Digital Literacy and Informal Learning Environments: An Introduction. *Learning, Media and Technology*, 38 (4). 355–367.

Mills, K., Bonsignore, E., Clegg, T., Ahn, J., Yip, J., Pauw, D., Cabrera, L., Hernly, K. and Pitt, C. (2018). Designing to Illuminate Children's Scientific Funds of Knowledge through Social Media Sharing. In: *Proceedings of the 17th ACM Conference on Interaction Design and Children*. Trondheim, Norway, 19–22 June 2018. 266–277.

Mills, K., Bonsignore, E., Clegg, T., Ahn, J., Yip, J., Pauw, D., Cabrera, L., Hernly, K. and Pitt, C. (2019) Connecting Children's Scientific Funds of Knowledge Shared on Social Media to Science Concepts. *International Journal of Child-Computer Interaction*, 21 (Sep). 54–64.

Moje, E.B., Ciechanowski, K.M., Kramer, K., Ellis, L. and Carrillo, R., & Collazo, T. (2004) Working toward Third Space in Content Area Literacy: An Examination of Everyday Funds of Knowledge and Discourse. *Reading Research Quarterly*, 39 (1). 38–70.

Moll, L.C., Cathy, A., Neff, D. and Gonzale, N. (1992) Funds of Knowledge for Teaching: Using a Qualitative Approach to Connect Homes and Classrooms. *Theory Into Practice*, 31 (2). 132–141.

NESA. (2017). New South Wales Educational Standards Authority [NESA], Science and technology syllabus K–6, 2017, available from: https://educationstandards.nsw.edu.au/wps/portal/nesa/k-10/learning-areas/science/science-and-technology-k-6-new-syllabus (21 January 2021).

Nishida, K. (1958) *Intelligibility and the Philosophy of Nothingness*.Tokyo, Japan: Maruzen.

Okoli, C. (2015) A Guide to Conducting a Standalone Systematic Literature Review. *Communications of the Association for Information Systems*, 37 (1). 879–910.

Papert, S. (1986) *Constructionism: A New Opportunity for Elementary Science Education*. Cambridge, USA: Massachusetts Institute of Technology, Media Laboratory, Epistemology and Learning Group.

Pei, L. and Nardi, B.A. (2019). We Did It Right, but It Was Still Wrong: Toward Assets-Based Design. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow, Scotland UK, 4–9 May 2019. alt07, 1–11.

Quayyum, F., Bueie, J., Vidal, J.C.T. and Jaccheri, L. (2020). Understanding Coding Activities for Teens: A Focus on School Teachers' Perspectives. In: *Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts*. London, United Kingdom, 21–24 June 2020. 187–192.

Read, J.C. and Bekker, M.M. (2011). The Nature of Child Computer Interaction. In: *Proceedings of HCI 2011 - 25th BCS Conference on Human Computer Interaction*.1–9.

Rembert, D.M., Mack, N.A. and Gilbert, J.E. (2019). Exploring the Needs and Interests of Fifth Graders for Personalized Math Word Problem Generation. In: *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. Boise, ID, USA, 12–15 June 2019. 592–597.

Roumelioti, E., Gennari, R., Matera, M., Melonio, A. and Rizvi, M. (2020). Towards Making Children Independent in Design. *Companion Publication of the 2020 ACM Designing Interactive Systems Conference*. Eindhoven, Netherlands, 6–10 July 2020. 227–232.

Salvi, V. (2017) Vulnerable Children as Change-Agents within Their Communities: An Educational Methodology Co-Designed in Namibia. *The Design Journal*, Routledge, 20 (sup1). S1485–S1502.

Saxe, G.B. (1991) *Culture and Cognitive Development: Studies in Mathematical Understanding*. Lawrence Erlbaum Associates.

Schepers, S., Dreessen, K. and Zaman, B. (2018) Rethinking Children's Roles in Participatory Design: The Child as a Process Designer. *International Journal of Child-Computer Interaction*, 16 (June). 47–54.

Schepers, S., Schoffelen, J., Zaman, B. and Dreessen, K. (2019) Children's Roles in Participatory Design Processes: Making the Role of Process Designer "Work". *Interaction Design and Architecture(S)*, 41. 87–108.

Scollon, R. and Scollon, S.W. (2004) *Nexus Analysis. Discourses and the Emerging Internet*. London, UK: Routledge.

Smith, R.C., Iversen, O.S. and Hjorth, M. (2015) Design Thinking for Digital Fabrication in Education. *International Journal of Child-Computer Interaction*, 5 (Sep). 20–28.

Södergren, A.C. and Van Mechelen, M. (2019). Towards a child-led design process A pilot study: When pre-schoolers' play becomes designing. In: *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. Boise, ID, USA, 12–15 June 2019. 629–634.

Soleimani, A., Green, K.E., Herro, D. and Walker, I.D. (2016). A Tangible, Story-Construction Process Employing Spatial, Computational-Thinking. In: *Proceedings of the The 15th International Conference on Interaction Design and Children*, Manchester, United Kingdom, 21-24 June 2016. 157–166.

Swidler, A. (1986) Culture in Action : Symbols and Strategies. *American Sociological Review*, 51 (2). 273–286.

Tondeur, J., Sinnaeve, I., van Houtte, M. and van Braak, J. (2010) Ict as Cultural Capital: The Relationship between Socioeconomic Status and the Computer-Use Profile of Young People. *New Media and Society*, 13 (1).151–168.

Troiano, G.M., Chen, Q., Alba, Á. V, Robles, G., Smith, G., Cassidy, M., Tucker-Raymond, E., Puttick, G. and Harteveld, C. (2020). Exploring How Game Genre in Student-Designed Games Influences Computational Thinking Development. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu, HI, USA, 25–30 April 2020. 1–17.

Troiano, G.M., Snodgrass, S., Argimak, E., Robles, G., Smith, G., Cassidy, M., Tucker-Raymond, E., Puttick, G. and Harteveld, C. (2019). Is My Game Ok Dr. Scratch?: Exploring Programming and Computational Thinking Development via Metrics in Student-Designed Serious Games for STEM. In: *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. Boise, ID, USA, 12-15 June 2019. 208–219.

Tuhkala, A., Wagner, M.L., Iversen, O.S. and Kärkkäinen, T. (2019) Technology Comprehension — Combining Computing, Design, and Societal Reflection as a National Subject. *International Journal of Child-Computer Interaction*, 20 (June). 54–63.

Weibert, A., Marshall, A., Aal, K., Schubert, K. and Rode, J.A. (2014). Sewing Interest in E-Textiles: Analyzing Making from a Gendered Perspective. In: *Proceedings of the 2014 Conference on Designing Interactive Systems*. Vancouver, BC, Canada, 21–25 June 2014. 15–24.

Weibert, A., Mouratidis, M., Khateb, R., Rüller, S., Hosak, M., Potka, S., Aal, K., and Wulf, V. (2017). Creating Environmental Awareness with Upcycling Making Activities. In: *Proceedings of the 2017 Conference on Interaction Design and Children*. Stanford, CA, USA, 27–30 June 2017. 286–291.

Wing, J.M. (2006) Computational Thinking. *Communications of the ACM*, 49 (3). 33–35.

Wong-Villacres, M., Disalvo, C., Kumar, N. and Disalvo, B. (2020). Culture in Action: Unpacking Capacities to Inform Assets-Based Design. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu, HI, USA, 25–30 April 2020. 1–14.

Yarosh, S., Radu, I., Hunter, S. and Rosenbaum, E. (2011). Examining Values: An Analysis of Nine Years of IDC Research. In: *Proceedings of the 10th International Conference on Interaction*

Design and Children. Ann Arbor, USA, 20-23 June 2011. 136–144.

Yip, J., Ahn, J., Clegg, T., Bonsignore, E., Pauw, D. and Gubbels, M. (2014). “It Helped Me Do My Science.” A Case of Designing Social Media Technologies for Children in Science Learning. In: *Proceedings of the 2014 Conference on Interaction Design and Children*. Aarhus, Denmark, 17–20 June 2014. 155–164.

Technologies Persona's Dilemma between Academics and practitioners - Ideation Guidelines for Co-designing Future Workshops

1st Author's name
Author's affiliation
Authors' address
Author's email address

2nd Author's name
Author's affiliation
Authors' address
Author's email address

3rd Author's name
Author's affiliation
Authors' address
Author's email address

Abstract. The introduction of Industry 4.0 interactive technologies and automated systems in complex organizations have imposed novel challenges and burdens on academics and industrial practitioners for developing systems that work for future workplaces. Developing such systems need sufficient knowledge and understanding of the trends and technological developments and viability from industry and academic experts before introducing them to the general population. Co-designing workshops with employees and users supported by various design tools can provide better ideation for designing future scenarios. We conducted a qualitative study to analyze academics' and industrial practitioners' views on a persona as a design tool during a conference workshop. These participants empirically test the co-creation of personas and find conceptual differences between the groups in their tool use. We propose guidelines on using personas for idea management in the co-design of future scenarios using pre and post-workshop surveys and workshop transcripts to code and clustered our findings. The conclusion is that considering the differences in academics' and industrial practitioners' perspectives, using the right design tools for ideation in a prepared environment for a combined team is substantial and can lead to designing positive experiences in future workplaces.

Future technologies, Personas, Co-designing workshops, Academics, Practitioners.

1. INTRODUCTION

Advanced machines can simultaneously perform multiple functions – a benefit that human beings cannot provide (Parasuraman et al., 2000). Industry 4.0 transition will bring more autonomous future technologies. Despite the numerous benefits of automation, it is essential to determine which functions should be automated and to what extent (Parasuraman et al., 2000). Therefore, there is a burden on interactive system designers for sound design work for its users. The analysis presented in this article has been developed to elicit different perspectives from academics and industry practitioners in understanding the nature and influence of using personas as a design tool in professional design work for future technologies. In this study, academics and industrial practitioners refer to individuals who engage in design activities at any level: the former with both academic and problem-solving motivations and the latter with an interest or responsibilities to find solutions to domain-specific problems.

In comparison, academics create more insights on selecting and evaluating design ideas as fundamental skills. In practice, the success of industry practitioners is based on experience rather than theoretical knowledge (Inie and Dalsgaard, 2020). However, despite disputes and differences between stakeholders' internal cultures or communication (Sarin and O'Connor, 2009), a common understanding and a shared vision to develop a new design must be facilitated to an efficient outcome.

1.1 Future Technologies

Predicting how the industries will develop or what future services will look like should be based on grounded foundations. The result can be sufficient knowledge about future workplace interactive technologies to recognize desirable or undesirable possibilities. The decision-makers can use this result as a broad understanding to avoid the negative consequences of using such future technologies (Jenkins et al., 2020).

In the design process, it is popular to use simple story studies (Jenkins et al., 2020; Kymalainen et al., 2016) based on trends and events to capture future possibilities. This can generate ideas by facilitating brainstorming sessions for more realistic scenarios or adopting 'blue sky thinking' to turn current trends and signals into the future experience (Jenkins et al., 2020). The result most probably impacts people's decision-making and increases creative capacity when considering the future.

1.2 Co-designing and selecting the right tool.

Co-design is used as a central approach in designing services in support of multi-disciplinary collaboration or future technologies. It may be defined as 'the creativity of designers and people not

trained in design working together in the design development process' (Sanders and Stappers, 2008). Co-designing artifacts with design tools is becoming increasingly popular for generating and managing ideas to establish a shared understanding and shared vision and goals in the early stages of an innovation process (Rygh and Clatworthy, 2019). However, using these approaches through these design tools to design new technologies is challenging and needs a facilitated environment to support the participant to ensure success.

Using tangible design tools can establish a shared 'language' through physical form when verbal communication fails due to professional terminology and misalignments between different professional working cultures (Jenkins et al., 2020; Kymalainen et al., 2016; Rygh and Clatworthy, 2019). On the one hand, it is vital to match the design tools and the people interacting with them in the co-designing sessions. Using these design tools is often developed on a trial-and-error basis, resulting in low engagement, trust, and interaction with the participating stakeholders (Rygh and Clatworthy, 2019). An example of such tools is the persona, a ubiquitous design tool in many studies and, not surprisingly, a widespread tangible tool in the industry. (Jansen et al., 2020; Nielsen, 2019; Tomlin, 2018)

A persona is a communicational tool typically used within User-Centred Design (Nielsen, 2019), which was introduced by Alan Cooper (Cooper, 2004) for the first time for mass-market software development. Personas in co-design projects begun to include users and others in either persona inceptions or assemblies or its deployment. Personas could provide a deeper understanding of improving the quality of work and the work environment for future scenarios (Cajander et al., 2015).

This paper asks the research question: What are the main differences in academics' and industrial practitioners' points of view on common design tools for ideation? What are the main requirements for preparing the environment for co-creation of personas and find conceptual differences between the groups in their tool use?

2. METHODOLOGY

We investigated a selection of design tools with particular attention to personas for co-designing (Simonsen, Jesper and Robertson, Toni, 2012) workshops for future scenarios. The selection criteria for these tools are based on suitability with the work domains of the designer, namely Industry 4.0, and automation, and convenience for the type, space, and duration of the workshops. We conducted a 5-hour workshop during a well-known HCI conference on co-designing personas for user experience (UX) and engagement in automation to gather this input. We also invited industry practitioners to join us to get their insights on the topic at hand. Due to COVID-19 outbreak, this workshop was run on an online platform virtually. The goals of the workshop, participants discussed:

- Presenting different views on using idea management workshops as a fundamental practice for future scenarios for industry practitioners and academics.
- Offering guidelines on using Personas as a design tool for idea management in the co-design of future scenarios.

2.1 Participant and Data collection process

The participants in this workshop were three industry practitioners from different companies, one doctoral student, three active researchers, and one post-doctorate researcher. The industry practitioners were a consultant for government projects, an innovation consultant with experience in innovation and user experience for a busy airport, and a business developer representing the UK-based start-up working with the high-ranking companies for the industry 4.0 transition.

We employed the Nominal Group Technique (Delbecq et al., 1975) for data collection during the workshop. This method is a structured method for group brainstorming that encourages contribution from everyone and quicker decision-making on the important relative issue, problem, or solutions. This technique helped all participants to have their contributions first and then discuss them with others. We can divide the workshop activities into three phases.

(i) **Introduction of the participants and the workshop goals.** Most participants presented their findings of the topic as part of their contribution, while others could ask questions and discuss their opinion on utilizing personas in the design process!

(ii) **Idea generation phase.** In this phase, we used a personas template (Nielsen, 2019) and invited the participants to co-design two personas for a defined scenario. They were discussions lead to questions on the situation, obstacles, domain context, users, and even the technology.

(iii) **Idea sharing phase.** Each participant described their notes on the scenario and raised the issues. The template allowed the

participant to cluster their ideas under each relevant heading.

After the activity, the facilitator sent a survey asking participants for insights at the end of the workshop. We clustered similar insights/findings under the initial categories.

3. WORKSHOP RESULT

The workshop's main activity focused on co-designing personas for a scenario introduced by the UK-based start-up focused on retail shopfloor robots in future supermarkets. During the Idea Generation phase, the workshop participants individually reported 50 different insights based on their previous works or research about using personas as a design tool. Subsequently, during the Idea Sharing phase, the participants discussed their different views on each insight. This number decreased to 21 as the number of duplicated insights was removed from the list. In this phase, participants benefited from each other's point of view and agreed on the number of the points raised as disagreement before. After this discussion, the final main categories were clustered into 9 insights (covered in Tables 1 and 2). Table 1 shows the main disagreement and table 2 is a finalized summary of the agreed insights on standard practices.

Table 1. Disagreed insights between the academics and industry practitioners on the designing personas.

Academics	Industry practitioners
Data driven persona vs proto personas	
Incorrect practices on personas by industry. 3 votes	Proto personas and reverse engineering as a common practice. 4 votes
Helpfulness of personas in designing future scenarios	
Co-designing for future scenarios better than the current ones. 6 votes	Not easy to create personas for new segments. 2 votes
The number of personas to design	
Businesses should prioritize the number of personas. 3 votes	No particular number to follow; the team will decide on this number. 4 votes
Type of data collection for personas	
Data for personas can come from different sources and forms. (various votes and ideas)	Using segmentation in busy domains is impossible, but alternative sources are available. (agreement on some of the alternative sources)

Table 2. Agreed insights between the academics and industry practitioners on co-designing personas.

Insights agreed
Personas can be used in iterative design and for testing the systems.
A.I. personas should consider future scenarios, specifically for technologies that have livelihood about their jobs and work alongside the human worker (e.g., Robots, A.I. engines).
Photo personas can transfer pre-conceptions about the users to the developers, or it can touch cultural sensitivity.
Data collection is not always accessible, but having a context, situation, obstacles, and a scenario is essential for co-designing personas.
A shared, engaging environment that people can design together is important for co-designing the personas and can increase the chance of practicable personas.

4.GUIDELINES ON USING PERSONAS FOR IDEATION FOR FUTURE SCENARIOS CO-DESIGNING WORKSHOPS.

We noticed that the ideation process could be sped up by providing a context, scenario, and better facilitation during the co-designing workshop. Previous studies cover how this process can be more comfortable and quicker for the participants (Inie and Dalsgaard, 2020; Rygh and Clatworthy, 2019). We add to this literature by presenting insights from our workshop.

Selecting the right design tool: Selecting design tools for co-design for future scenarios should be based on the workshop medium (i.e., online or face-to-face) and participants' familiarity with the design process and tools (i.e., professional, expert, typical end-users). Participants select the tools they are more confident with and tools they may not need specific knowledge or literacy for working with (e.g., sticky notes). Adding more tangible elements (Rygh and Clatworthy, 2019) such as pre-structured cards and easy-to-use collaborative tools can always be beneficial.

Preparing the environment to ideate:

Whether it is a face-to-face or virtual co-design workshop, participants should understand the scenario, products, and future system's domain and environment. For example, using employees as participants facilitate familiarity with the work domain. However, preparing all participants with brainstorming, group discussion, etc can provide them with the exemplary scenario and situation to generate and manage their ideas. Nevertheless, during our workshop, we have noticed that we do not only need collaborative design tools, but we also need facilitation with a communication system, ideally recording the sessions and discussions.

Refining the tools for future scenarios:

We used a persona template (Nielsen, 2019), which has a part for the scenario, unlike many common persona templates widely available. In particular, this section was helpful in a virtual co-design workshop as it provides a platform to test the persona in the given scenario. Service design guides introduce more useful design tools than personas with more actors to look at different levels in the user journey (i.e., user stories and blueprint), thus having a new investigation on designing more contextual personas with the new requirements is needed (Cajander et al., 2015). We used an approach towards a scenario and a persona which brought more interaction into the workshop. Consequently, it brought up more issues to deal with (e.g., trust, hierarchies, and persona's limitations) rather than using only one of them. We agree that adding the two sections for the context data and external and internal sources to build the personas is necessary, considering more online collaborations to come.

Personas for A.I.:

Industry practitioners supported the idea of creating 'Personas for A.I.' in the introduction phase, which was challenged by some academics at the beginning of the workshop. Practitioners emphasized the importance of A.I. behavior in different domains. For example, in airports, A.I. systems are associated with a level of risk as they will take away some controls from human counterparts. Therefore, there is a need for generating a persona for such systems, one of the participants said. Likewise, UK-based start-up's operation team has created different robot profiles in the airport and retail's environment based on A.I. behaviour. Ultimately, one of the consultants suggested looking into a new design and phrase for personas for technology to be called 'Techsona' (Technology Persona).

Instruction for facilitator during the activities:

The facilitator's role is significant in such co-design workshops to help and manage the ideation process. Knowing the main topics to cover or being open about the new topic can create more real personas, such as raising privacy or GDPR topics while co-designing personas for the retail shopfloor robot's scenario in this workshop. As agreed by participants, facilitators should be carefully selected and trained to run the co-design sessions. They need to follow instructions and use carefully selected activities. A participant who played the facilitator role in the workshop, after the co-design activity, suggested that facilitators should supervise the outcome and tidy up the outcome of the idea management activity. We also suggest collaborations between academics and industry for co-designing future technologies. For instance, the facilitator can be selected from academia working with the industry. To make the personas or any other design tools in line with the organization's UX, providing a short, easy-to-understand template should be considered. One of the academics in the workshop believed that the UX\AI goals could help shape the use-cases if facilitated with a proper exercise.

Importance of scenarios:

In line with the environment preparation and facilitator's instructions, it is essential to plan for a scenario for every co-design workshop for future scenarios. This can help the participant access context and a road map to look forward and share more relevant information. One academic believed personas should be tested on scenarios and refined again and again for an optimal outcome closer to reality for future technologies. The industry practitioners in the team supported his idea strongly.

Data gathering for persona designing:

There is a need to use data-driven design tools to co-design for future scenarios. Accessibility to this data in the templates, collaborative platforms, or any other forms can provide better context and understanding for the participants before using these tools.

5. CONCLUSION

We explored empirical differences in academics' and industry practitioners' perspectives in using persona as a design tool to ideate future scenarios.

We analyzed the collected data from a conference workshop. Using participant's point of views, we recommend a guideline for running co-design workshops. Our analysis indicated that while these two groups may, on the behavioural level, use common design tools in the same fashion, there are important differences between them in the thinking, ideation, prototyping, and overall design process. We can argue that (1) there is a need to involve both academics and industry practitioners in co-design ideation workshops for emergent future systems, (2) design tools used should be linked explicitly to specific contexts, scenarios, or situations to provide decision information relevant to the specific domain and environment, (3) using data-driven design tools based on the current trends and events may facilitate consensus about a design reality. Also, the facilitators of co-design ideation workshops have a critical role in leading the various designers and the overall session to an optimal outcome.

6. REFERENCES

Cajander, Å., Larusdottir, M., Eriksson, E., Nauwerck, G., 2015. Contextual Personas as a Method for Understanding Digital Work Environments, in: Abdelnour Nocera, J., Barricelli, B.R., Lopes, A., Campos, P., Clemmensen, T. (Eds.), Human Work Interaction Design. Work Analysis and Interaction Design Methods for Pervasive and Smart Workplaces, IFIP Advances in Information and Communication Technology. Springer International Publishing, Cham, pp. 141–152. https://doi.org/10.1007/978-3-319-27048-7_10

Cooper, A., 2004. The Inmates are Running the Asylum: Why High-tech Products Drive Us Crazy and How to Restore the Sanity., 1st Edition. ed. Sams.

Delbecq, A.L.., Van De Ven, A.H., Gustafson, D.H., 1975. Group techniques for program planning: a guide to nominal group and delphi processes, Management applications series. Scott, Foresman.

Inie, N., Dalsgaard, P., 2020. How Interaction Designers Use Tools to Manage Ideas. ACM Trans. Comput.-Hum. Interact. 27, 7:1-7:26. https://doi.org/10.1145/3365104

Jansen, B.J., Salminen, J.O., Jung, S.-G., 2020. Data-Driven Personas for Enhanced User Understanding: Combining Empathy with Rationality for Better Insights to Analytics. Data and Information Management 4, 1–17. https://doi.org/10.2478/dim-2020-0005

Jenkins, T., Boer, L., Brigitta Busboom, J., Simonsen, I. Østby, 2020. The Future Supermarket: A Case Study of Ethnographic Experiential Futures, in: Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society. Presented at the NordiCHI '20: Shaping Experiences, Shaping Society, ACM, Tallinn Estonia, pp. 1–13. https://doi.org/10.1145/3419249.3420130

Kymalainen, T., Kaasinen, E., Aikala, M., Hakulinen, J., Heimonen, T., Paunonen, H., Ruotsalainen, J., Lehtikunnas, L., Mannonen, P., 2016. Evaluating Future Automation Work in Process Plants with an Experience-Driven Science Fiction Prototype, in: 2016 12th International Conference on Intelligent Environments (I.E.). Presented at the 2016 12th International Conference on Intelligent Environments (I.E.), IEEE, London, United Kingdom, pp. 54–61. https://doi.org/10.1109/IE.2016.17

Nielsen, L., 2019. Personas - User Focused Design, Human–Computer Interaction Series. Springer London, London. https://doi.org/10.1007/978-1-4471-7427-1

Rygh, K., Clatworthy, S., 2019. The Use of Tangible Tools as a Means to Support Co-design During Service Design Innovation Projects in Healthcare, in: Pfannstiel, M.A., Rasche, C. (Eds.), Service Design and Service Thinking in Healthcare and Hospital Management. Springer International Publishing, Cham, pp. 93–115. https://doi.org/10.1007/978-3-030-00749-2_7

Sanders, E.B.-N., Stappers, P.J., 2008. Co-creation and the new landscapes of design. CoDesign 4, 5–18. https://doi.org/10.1080/15710880701875068

Sarin, S., O'Connor, G.C., 2009. First among Equals: The Effect of Team Leader Characteristics on the Internal Dynamics of Cross-Functional Product Development Teams. Journal of Product Innovation Management 26, 188–205. https://doi.org/10.1111/j.1540-5885.2009.00345.x

Simonsen, Jesper and Robertson, Toni, 2012. Routledge international handbook of participatory design. Routledge.

Tan, W., Boy, G.A., 2018. Tablet-Based Information System for Commercial Aircraft: Onboard Context-Sensitive Information System (OCSIS), in: Harris, D. (Ed.), Engineering Psychology and Cognitive Ergonomics, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 701–712. https://doi.org/10.1007/978-3-319-91122-9_55

Using Card Sorting to Design Faceted Navigation Structures

Ed de Quincey
School of Computing and Mathematics
Keele University
e.de.quincey@keele.ac.uk

James Mitchell
School of Computing and Mathematics
Keele University
j.a.mitchell@keele.ac.uk

This paper describes how a variation of card sorting- 'repeated single-criterion sorting', can be applied to the information architecture design of digital music services. 52 respondents were asked to sort, using their own choice of criteria, 12 popular songs using an online card sorting tool. Once respondents had chosen a construct for a particular sort, e.g. "Genre", they placed each card into a named category, e.g. "Rock", "Pop", and were encouraged to repeat this process until they could think of no more constructs. High levels of agreement were found for a small number of constructs such as "genre", "gender", and "speed of song" but the remaining constructs were individual to each respondent, e.g. "songs that make me cry". The results highlighted differences with current approaches to music categorisation, as well as the potential for repeated single-criterion sorting to be used to design faceted navigation structures.

Requirements Elicitation, Card Sorting, Information Architecture, Faceted Navigation, User-Centered Design (UCD), User Experience (UXD, UX).

1. INTRODUCTION

User experience design (UXD) and usability evaluation are supported by a range of tools and techniques. However, a number of these methods utilise attributes predetermined by the researcher, such as heuristic assessment (Nielsen & Molich, 1990) and cognitive walkthroughs (Wharton *et al*., 1994), which may not be relevant to the intended audience. To tackle this, methods that elicit criteria from the users themselves, such as think-aloud protocol (Nielsen, 1992), card sorts (Rugg & McGeorge, 1992), and laddering (Gammack, 1987) have been developed.

A popularised version of card sorting, termed "all-in-one" sorting, has become one of the standard techniques in the User-Centered Design (UCD) process (Usability.gov. N.D). "All-in-one" sorting improves "findability" within a system and has been successfully used to determine the Information Architecture (IA) of websites (Spencer & Warfel, 2009) (e.g. Frederickson-Mele, 1997; Tullis, 2003; Tullis & Wood, 2004) and software menu systems (Tullis, 1985). In addition, a variation of card sorting, 'repeated single-criterion sorting', has been highlighted (Maiden, 2009) for its potential in requirements elicitation, but few empirical studies investigate how this variation of card sorting relates to UXD and how it can be integrated into the UCD process.

In order to address this, this study has applied repeated single-criterion sorting to the problem of music categorisation and the design of digital music services.

This paper is structured as follows. Section 2 gives an overview of different card sorting techniques and work related to music categorisation. Section 3 outlines the method used, respondents and materials. Section 4 presents the results and Section 5 the analysis of these results. Section 6 discusses the implications of the study and Section 7 presents overall conclusions.

2. BACKGROUND

This section describes the different card sorting methods and current academic and commercial approaches to music categorisation.

2.1 Card Sorting

All-in-one sorting
This category of sorting covers a wide range of methodologies. The consistent feature is that the respondent(s) sorts the set of entities once. The respondent(s) are given a set of cards which represent products, pages, or functionality, within the site or application. The number of cards depends on the entity (Courage & Baxter, 2005). Up to five hundred cards however have been sorted in previous studies (Tullis, 1985). The respondent sorts the entities into a set of pre-defined categories, or categories of the respondent's choosing and name them. In the case of closed sorting the category that each entity has been sorted into is recorded; for open sorting the names of the categories are recorded along with the category position of each entity. Open card sorting "is useful as input to information structures in new or existing sites and products" whereas "closed sorting is useful when adding new content to an existing structure, or for gaining additional feedback after an open card sort" (Spencer & Warfel, 2009). The advantages include savings in time and cost, and as it is user-centered, it is therefore not as susceptible to "gut-feel" biases (Spencer & Warfel, 2009). Disadvantages include that the methodology is content-centric and fails to take in to account the users' task i.e. how the user interacts with content on a site and that analysis can be time-consuming-especially with large numbers of cards and/or respondents (Spencer & Warfel, 2009).

"All-in-one" sorting is generally used to determine IA and uses variations of cluster analysis to determine the navigational structure of a website or application. It has been used successfully in the design of large-scale websites e.g. the Google AdWords Help Center (Nakhimovsky et al., 2006) and is now an integral part of the analysis and design stages of a UCD lifecycle (Bevan, 2003).

Repeated single-criterion sorts
Repeated single-criterion sorting (or open card-sorting) involves asking the respondent to sort entities into groups of their own choosing; then to sort again, using a different criterion of their own choosing, until they run out of criteria. For example, if the entities are shoes, "colour" might be the first criterion, with categories such as "brown", "black" and "white". They may be further sorted by "material" as the criterion, and into categories such as "leather" and "canvas". Empirical research has found no statistical difference between the types of criteria and categories elicited when using different types of entity (Rugg et al., 1992).

This approach works well with nominal categories, and typically elicits group names and criteria consisting of short phrases. This method is described by (Gammack, 1987) and is described in detail in a tutorial paper by (Rugg & McGeorge, 1997). The technique has been applied to a wide range of topics, including web page quality metrics (Upchurch *et al*., 2001); quantification of copyright infringement (Martine & Rugg, 2005), and assessment of differences between expert and student programmers (Sanders et al, 2005). It is supported by a range of statistical analyses, including co-occurrence matrices (Martine & Rugg,

2005) and minimum edit distances (Deibel *et al*., 2005).

Rugg & McGeorge (1997) recommend the use of repeated single-criterion sorting for requirements elicitation due to its flexibility and a stronger grounding with the relevant theoretical foundation i.e. Personal Construct Theory (PCT). Maiden (2009) has also highlighted the value of this type of sort for requirements elicitation for similar reasons. Repeated single-criterion sorting has been used to study perceptions, such as identifying the features of web pages that users are interested in (Upchurch et al., 2001). However, this method is not formally aligned with the UCD process in the same way that "all-in-one" sorting is.

Music Categorisation

The ease of purchasing and streaming music online and the shift towards storing and organising music digitally has shaped music preference behaviour (Greasley & Lamont, 2006). Studies investigating people's use of music have indicated that people listen to music for specific reasons and that their motivations for listening to music depends on context (DeNora, 2000; North et al., 2000; Sloboda et al. 2001). However, the ubiquitous nature and ease of access to music presents problems: How to organise music so that it is accessible, and convenient? How to discover more songs, similar to those they enjoy?

Musical genre is a widely used standard for categorising music (Aucouturier & Pachet, 2003; Pachet & Cazaly, 2000) and often the preferred technique. However, the definition of a music genre is subjective since it is influenced by extrinsic factors (Lippens, 2004; Aucouturier & Pachet, 2003). This leads to undefined boundaries of genres and as a consequence there is a lack of a precise method of classifying music to genres. Online stores such as iTunes categorise music by standard music industry decided taxonomies (similar layout of traditional bricks and mortar retail stores) with genre being the primary method for users of the software to find songs that they like. The user has to have a definite idea of what genres they like, what genre a song fits into and for this to match with the categorisation used by the online store.

One of the most commercially popular attempts at improving musical classification with the intention of creating automated playlists for streaming radio is Pandora, based upon the Music Genome Project (Joyce, J., 2006). The Music Genome project attempts to describe music with vectors consisting of hundreds of genes or musical attributes describing each song (McKay, 2010). It is unclear what the complete list of attributes are, but a partial list (that is now not publicly available) suggested that the following are included: Structures; Roots; Tonality; Instrumentation; Feel; Musical qualities; Leanings/styling; Recording techniques; Influence; Instruments; Lyrical content and Vocals.

A song is represented by a vector of these attributes, with up to five hundred genes/attributes forming the vector. Each attribute/gene is assigned a number between 1 and 5 in half-integer increments (see Music Genome Project US Patent: No. 7,003,515). Given a single song or a group of songs, a distance measure is then calculated from this vector to produce a list of related songs. Though the retrieval process is automated, the scaling and classification of the songs is an entirely manual, subjective process.

There have been attempts at creating flat taxonomies of music using folksonomies or tagging e.g. Last.fm, which utilise the user generated attributes to categorise songs, increasing the number of attributes that can then be used to describe a song and potentially produce a richer categorisation schema. These rely solely on the individual's perceptions of a song and depend on users using the same tag or set of tags for the same songs if they can be then used to recommend similar songs.

For the majority of digital music services, there is a reliance on genre-based systems which are reliant on inconsistent, intrinsic features of a song. There is little research into appropriate feature sets for classifying different types of song, and the relationship between objective and subjective attributes that users classify songs by. There is also little exploration into which of these features can be supported by current technologies or by extending pre-existing functionality as opposed to creating entirely new systems. Card sorting offers a potential solution to these problems as it has been previously used to determine user perceptions of a range of media (e.g. Upchurch et al., 2001) and also similarity measures using a range of user identified attributes (Martine & Rugg, 2005).

The following sections describe the use of repeated single-criterion card sorting to identify users' perceptions of a range of popular songs.

3. METHODS

As part of a related study (de Quincey, 2010), an online card sorting tool has been developed. The application includes functionality to support a range of multimedia such as pictures, music and videos along with analysis techniques such as co-occurrence matrices.

Figure 1 shows the sorting interface where respondents are presented with the entities on the left-hand side (in this case artist names) and are asked to input the sort criteria (in this case "Gender of Artist") and sort the entities into groups of their choosing (in this case the user has created 2 groups,

"Male" and "Female"). Users can then drag the entity into the appropriate group (see Figure 2). For multimedia entities, users double-click the card to view/hear the video/song.

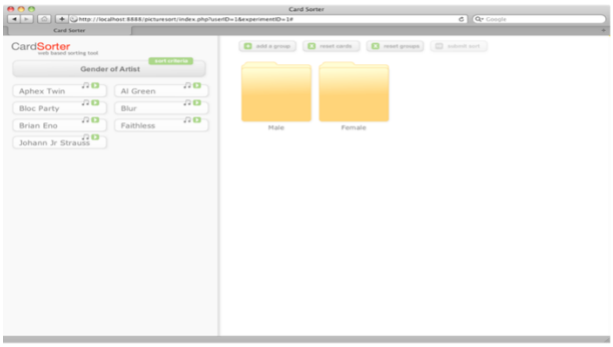


Figure 1: *Screenshot of online card sorting tool*

Figure 2*: Screenshot showing sorting behaviour*

Once the entities are sorted, the user is prompted to either perform another sort using a different criterion or to end the sort. The results are recorded automatically in a database

Respondents were asked to use the card sorting tool to sort, using their own choice of criteria, a number of popular songs. Once respondents had chosen a criterion for a particular sort, they placed each card into a named group and were encouraged to repeat this process until they could think of no more criteria. The researcher used a dyadic elicitation technique which involved playing two random music clips and asking the respondent whether they could think of any differences between the two songs which could form another sort criterion. The sessions were carried out under controlled conditions in the same room, using the same computer to remove unforeseen technical issues that may occur e.g. user not able to hear songs, slow internet connection etc. The sessions were undertaken between 2006 and 2007.

3.1 Respondents

There were a total of 52 respondents, 42 from the School of Psychology student pool and 10 from within the School of Computing and Mathematics. 32 females and 14 males participated (plus 6 participants who did not provide information on their gender). The participants' age range was 18 years to 37 years.

3.2 Materials

Twelve popular songs were chosen for the entities by researchers in the School of Psychology to complement research that was being undertaken by (Greasley & Lamont, 2006). Songs were from popular artists and represented a range of genre (see Table 1).

Table 14. *Songs and artists used as entities for sorting*

ID	Artist - Song
1	*Coldplay - Yellow*
2	*Eminem – Without Me*
3	*Misteeq – Why?*
4	*Rage Against the Machine - Wake Up*
5	*Maroon 5 - This is Love*
6	*UB40 - Red Red Wine*
7	*De La Soul - Three*
8	*Hard Fi - Living for the Weekend*
9	*Madonna – Hung Up*
10	*Chemical Brothers - Galvanise*
11	*Tracy Chapman – Fast Car*
12	*Mary J Blige – Family Affair*

Each song was cropped to the first thirty seconds of the song. The screen representation of each song (see Figure 1) that users double click to play was labelled with an arbitrary number between one and twelve. Using the title or artist as the card label was considered but this may have prompted criteria related to the song title or the artist, not the song itself.

4. RESULTS

This section presents the results from the 52 card sorting sessions, outlining the constructs and categories used and their distribution between respondents.

4.1 Number of constructs and categories

A total number of 295 constructs[2] were elicited from 52 respondents. The number of constructs per session ranged from 2 to 11. Respondents used between 2 and 9 categories for each sort with the majority of sorts comprising of dyadic (2 categories used) and triadic (3 categories used) sorts.

4.2 Commonality of constructs

From the 295 constructs, there was direct verbatim agreement i.e. two or more respondents using the exact same phrase, for 28 constructs. The most

[2] Criteria elicited during card sorting sessions of this type are normally known as constructs due to the link with PCT.

frequently used verbatim constructs are shown in Table 2. When scrutinising the verbatim constructs it was apparent that different respondents used different words for similar constructs e.g. “Music Type” and “Type of music”, “Tempo” and “Speed”. In line with previous research (e.g. Gerrard & Dickinson, 2005), an independent judge was used to group the constructs into superordinate constructs, giving an indication of commonality between respondents.

Table 15. *Commonly used verbatim constructs*

Verbatim Construct	Number of Subjects
Genre	23
Tempo	12
Gender of singer	10
Gender	9
Speed	7
Type of music	7
Artist(s)	5
Music type(s)	5
Style	5
Audience	4
Era	4
Group or solo	4

The judge was given a set of standard instructions and a full set of results including the constructs, category names and card groupings. A number of previous studies provided a list of grouped verbatim constructs to the independent judge. As identified in a previous study (de Quincey, 2010) this should remove constructs that use the same words but are related to different attributes (category names).

Table 16. *Commonly used superordinate constructs*

Superordinate Construct	No of constructs	% of Respondents
Genre of Music	45	88
Gender of Artist	34	67
Speed of song	27	53
Solo or group	18	35
Year music produced/released	16	31
Likeability of song	13	25
Main instrument	9	18
Audience	8	16
Emotion	7	14
Nationality of artist	7	14

When grouped into superordinate constructs, the number of constructs was reduced from 289 to 78. Table 3 shows the number of constructs included for each superordinate construct and the percentage of respondents that they were elicited from for the top 10 most used constructs. Following this grouping, agreement was found amongst respondents for 26 superordinates out of the 78 e.g. 88% of respondents used the “Genre of Music” as a construct and 67% of respondents used “Gender of Artist”. This shows a high level of commonality for a small number of the constructs with 52 constructs out of the 78 being generated by single users. Examples of these unique constructs included “Volume of drums”, “Complexity of music”, “Make you sad”, “Is it relaxing” and “Music to work to”.

4.3 Distribution of Items

As described by Martine & Rugg (2005), card sorts data can be used to produce co-occurrence matrices that give an indication of similarity between the entities represented by the cards and the distribution of entities for similar constructs. The matrix is produced by summing the number of times one card appears in the same category as another card for all respondents and criteria.

	1	2	3	4	5	6	7	8	9	10	11	12
1		85	50	119	169	150	110	151	60	92	151	65
2			115	102	104	87	119	118	129	167	63	134
3				58	80	61	79	97	181	132	61	174
4					114	104	121	147	52	116	89	65
5						112	130	176	108	109	94	92
6							127	95	49	97	155	76
7								123	76	136	98	93
8									100	131	81	83
9										116	75	146
10											61	126
11												79
12												

Figure 3. *Co-occurrence matrix for the 12 songs*

Figure 3 shows the matrix for the twelve songs. With the total number of sorts being 289, the maximum number of times for two cards to be placed in the same category is also 289. The highest amount of co-occurrence was 181 (for songs 9 and 3). The lowest of amount of co-occurrence was 49 (for songs 9 and 6).

To determine levels of agreement within superordinates i.e. whether respondents have placed the same songs in the same groups for similar criteria, co-occurrence matrices were calculated for the sorts where a number of respondents had used similar criteria as determined by the independent judge. The matrices in Figures 4 and 5 show the percentage of times that two songs were placed in the same group to allow for comparison between matrices. For example, in Figure 2, songs 1 and 4 were placed in the same group 94% of the time for sorts related to “Gender”.

For these matrices, if respondents agreed on the criterion, the matrix would at best case contain a selection of very high numbers (close to 100%) and very low numbers (close to 0%). This would indicate that certain songs would always be placed in the same group for that criterion and others would never

be placed in the same group. Plotting the histogram of this data should demonstrate bimodal distribution. For the superordinate "Gender" as shown in Figure 5, this is mostly the case. For example, songs 1 and 2 are placed in the same group 91% percent of the time when the criterion for the sort is related to "Gender" but songs 1 and 3 are never placed in the same group.

	1	2	3	4	5	6	7	8	9	10	11	12
1		91	0	94	94	94	94	94	0	94	41	0
2			0	88	91	94	94	91	3	94	38	3
3				0	0	0	0	3	97	0	53	97
4					91	88	91	91	0	91	35	0
5						88	94	94	0	94	35	0
6							94	88	0	94	44	0
7								94	0	100	41	0
8									3	94	41	3
9										0	53	100
10											41	0
11												53
12												

Figure 4. *Co-occurrence matrix for "Gender" (values represented as %)*

When respondents were sorting based on gender, then male vocalists would always appear in the same group and never appear with songs with female vocalists and vice versa. For the majority of songs this was the case, but song 11 was placed with most of the songs between 35% and 53% of the time. It seems that respondents were unable to consistently group the song into a specific gender. This may be due to the vocalist, Tracy Chapman, having an undeterminable voice.

For the other superordinates levels of agreement are less consistent. "Genre" related constructs were used by 88% of respondents, but the level of agreement between the respondents indicated by the matrix shown below are low. Songs 5 and 8 were sorted into the same group 60% of the time (which was the highest) but the majority of songs had low levels of co-occurrence.

	1	2	3	4	5	6	7	8	9	10	11	12
1		0	2	20	44	18	13	53	20	2	40	2
2			16	2	0	2	22	0	2	33	2	38
3				4	7	7	11	9	31	27	9	53
4					7	9	9	24	2	4	16	0
5						11	22	60	42	9	24	7
6							24	7	11	16	40	2
7								20	11	36	29	11
8									11	9	24	4
9										27	11	13
10											16	31
11												0
12												

Figure 5: Co-occurrence matrix for "Genre" (values represented as %)

The matrix for "Solo or Group" showed higher levels of agreement with a number of songs (5 and 1, 9 and 2, 2 and 11, 8 and 1) co-occurring over 90% of the time. The matrix for "Speed of Song" highlighted some songs that are perceived to be similar (9 and 3, 11 and 1) but the majority show low levels of co-occurrence. For this construct it would be expected that songs would form clear groups based on the time signature that they were written in e.g. 3/4 time compared to 4/4 time but that does not seem to be the case. Matrices for the constructs "Likeability of Song" and "Year Produced" show similar distributions to "Speed of Song" with some songs with high levels of similarity, but the majority having low levels of co-occurrence.

5. ANALYSIS

5.1 Constructs used

Respondents generated a large range of constructs (2 to 11) and categories (2 to 9). The large number of constructs generated by some respondents suggests that they have expert knowledge in the domain (Rugg & McGeorge, 1997). This might be expected from the student population that the respondents were recruited from as some were studying music or involved with university orchestras etc.

High commonality (>50%) was found for a small number of superordinate constructs such as "genre", "gender" and "speed of song" but the remaining 75 constructs showed little agreement. This gives an indication that once "genre" and "gender" have been used, further constructs are individual to each respondent. Of these remaining constructs, there is a mixture of objective, such as "age of artist" and subjective criteria such as "would I pay to see them in concert".

5.2 Sorting behavior

There was little agreement between respondents in how they sorted the songs into categories for all criteria. The maximum number of times that two songs were placed in the same group was 181 (for songs 9 and 3) out of a possible 289 (62% of the time). This may be due to the songs being entirely different, different perceptions of the group a song fits into, or respondents using different criteria. The co-occurrence matrices demonstrate a range in the levels of agreement between respondents when sorting using the same/similar criteria.

Genre: Almost all of the respondents used "genre" as a criterion but there was little agreement between which songs fitted into the same genres or what those genres should be called. This indicates that due to "genre" being the default index method in music retail, people use what they are accustomed

to despite there being little agreement into which categories particular songs fall into.

Gender: When sorting using "Gender" as the criterion, respondents were consistent with the majority of songs, finding it easy to determine the gender of the vocalist in all but one case, but it is sometimes unclear as to what "gender" refers to. For the majority of constructs in this study the gender refers to the gender of the vocalist, but the distribution of songs may become inconsistent if there is more than one vocalist for a song.

Solo or Group: Surprising the co-occurrence levels for "solo or group" were again low, especially considering that the criterion is highly objective. This could indicate a respondent's lack of knowledge of the song artists or the inability to determine from a thirty second clip whether it is a solo or group artist.

Speed of Song: The co-occurrence levels for "speed of song" were also low indicating that respondents are inconsistent in their perceptions of what constitutes speed, even though the majority used the word "tempo" as the criterion name and "fast", "slow" and "medium" as categories. This echoes research by Scheirer (1998).

5.3 Summary

Some agreement was found in the criteria used for sorting but there were many constructs that were unique. Within the categories, agreement was only found for the "gender" construct, with some agreement for certain songs when using certain criteria. The differing levels of agreement regarding song categorisation have implications for digital music services that are described in the following section.

6. DISCUSSION

6.1 Implications for Digital Music Services

Having identified that there is some level of agreement between respondents for the criteria used to sort music, one possible use of these constructs would be to then integrate them into music library/streaming software such as iTunes, to improve the ability of users to organise and navigate music libraries. The following table (Table 4) details constructs that could potentially be used as attributes for categorising music that are already utilised by iTunes, Spotify or the ID3 specification. ID3 frames are a popular "audio file data tagging format" that are used by a number of popular music players (O'Neill, 2013). An ID3 tag is a data container within an MP3 file that is stored in a predefined format allowing users and artists/vendors to encode additional information into an MP3 file such as text or pictures. Currently iTunes, and similar software, do not use all of the frames that are defined within the standard and therefore specialised ID3 tag editors have to be used to edit the majority of the ID3 frames (although iTunes does allow the user to edit certain fields and automatically populates them via the iTunes store or via the Internet).

Table 4. *Constructs used in Digital Music Services*

Construct	iTunes	Spotify	ID3
Genre of Music	*Yes*	*Yes*	*Yes*
Speed of song	*Yes "BPM "*	*No*	*Yes: "Exact tempo codes"*
Year music produced/ released	*Yes*	*Partial: "Decades"*	*Yes "Date/Year of recording"*
Likeability of song	*Yes "iTunes Rating"*	*Partial: "Add to My Music"*	*Yes "Pupularimeter"*
Emotion	*No*	*Yes: "Moods"*	*No*
Place to listen to music	*No*	*Yes: "Focus", "Travel", "Dinner", "Sleep", "Workout"*	*No*
Chart position	*Yes "iTunes Chart"*	*Yes "Charts"*	*No*
Familiarity with song	*Yes "Play Count"*	*Partial "Your Music"*	*Yes: "Play counter"*
Mainstream or alternative	*No*	*Partial "Genres"*	*No*
Popularity of music	*Yes "iTunes Chart"*	*Yes "Charts", "Plays" and "Trending"*	*Yes "Pupularimeter"*

This table illustrates that of the 78 constructs, only 14 are utilised for organising or finding music. Spotify, has recently started to support more subjective ways of browsing music such as "Emotion" and "Place to listen to music" via their own curated playlists in the "Genres & Moods" area of their desktop application.

From the commonly used constructs shown in Table 3, "Gender of Artist", "Solo or group", "Main instrument", "Audience" and "Nationality of artist" are not currently supported by either iTunes, Spotify or ID3 tags.

The majority of the remaining unsupported constructs are highly subjective e.g. "Music to work to", "Times to listen to" suggests that automation of these parameters may be unrealistic and it would therefore be more practical to include functionality within the software that allowed users the freedom to sort and define songs with relevant constructs and attributes. Playlists provide this functionality to a

certain extent but specific ID3 tags for certain attributes that could be saved within the file (as opposed to within the software such as Playlists) would be preferred due to the potential for standardising and sharing this information, particularly for objective attributes such as “Gender of Artist” and “Main instrument”.

6.2 Automation of constructs/categorization

With large collections of music, manual curation of songs and playlists is non-trivial and automated methods of categorisation are appealing.

Over 50% of the respondents from this study used “Speed of song” as a criterion, suggesting that tempo is a widespread construct in perception of music and would be a suitable attribute to include in automated music retrieval and classification systems. Tempo is one of the basic attributes of music and has been used previously as a parameter for automatic information retrieval (Scheirer, 1998). The current support in music players relies on the artist/user manually including/editing the relevant “BPM (beats per minute)” ID3 frame or using software such as Media Center 9 to try and automatically detect the BPM (anecdotal evidence however suggests that this method is inaccurate). Automated rhythm and frequency methods have been previously used to identify the tempo and beat of a song and these methods have also been compared to people’s perceptions of tempo (e.g. Lippens et al., 2004).

Respondents from this experiment used the following criteria related to “speed of song”: Song speed, Speed, Slow and fast, BPS, Pace, Type of dance, Tempo, Music speed, Fast paced and Beat speed. The co-occurrence matrix for sorts related to “speed of song” indicated that there was a high level of disagreement between respondents even though they used similar category names e.g. “slow”, "fast”, “quick” etc. This indicates that even if an automated method for determining tempo is used, users themselves are using different attributes or measures for what constitutes the speed of a song.

6.3 Folksonomies and tagging systems

There has been increased interest in the idea of using folksonomies and tagging as a way of categorising and exploring music. A number of studies into the use of tagging and the related field of social bookmarking (Kipp, 2007; Kipp & Campbell, 2006; Golder & Huberman, 2005) suggest that tagging and bookmarking share similar features to more traditional indexing systems (Kipp & Campbell, 2006) but also contain extra dimensions such as tags related to time e.g. “toread” and task or users’ emotional responses to a document e.g. “cool”, which conventional indexing systems and approaches do not support (Kipp & Campbell, 2006).

Last.fm is a website that builds profiles of musical listening habits and also allows users to tag songs and artists with descriptive words or phrases. Comparing the constructs used by the respondents in this study with the tags generated by the users of last.fm highlights some interesting similarities and differences. The majority of tags used currently on last.fm (see Figure 19 in (de Quincey, 2010)) are genre based and are similar to the most common criteria used by respondents in this study e.g. “alternative”, “classis rock”, “electronic” etc.. There are also tags such as “female vocalists” and “male vocalists” that refer to the “gender of artist”, temporal tags e.g. “00s”, “80s” etc. that are similar to the “year music produced/released” and “favorites” “favourite”, “favourites” are similar to the “likeability of songs” construct.

It is interesting that “seen live” and “albums I own” are popular tags on last.fm but were not commonly used by respondents in this study (no respondents used “seen live” as a criterion and only one used a criterion related to “ownership”). Another point of interest is that the constructs “speed of song” and “solo or group” do not have equivalent last.fm tags.

6.4 Implications for Faceted Navigation

Amazon and Google make use of faceted navigation structures to allow users to further filter search results. The use of these structures can be linked to Facet Theory, originally devised as an improved way for categorising and indexing books by Ranganathan (Ranganathan, 1962). Generally, in these systems, once a user has performed a standard keyword search, as well as seeing the list of returned results, they are also given the option of searching/filtering within those results by various facets. This approach is often called “guided navigation” and although the term facet is not explicitly used, it is clear that providing users with options to search by format e.g. video, academic resources, images etc. or by geographical location or by time e.g. latest, past 24 hours, past year etc. the resources or products are being categorised by various facets.

Ranganathan’s approach to facets, deriving them systematically using Canons, Postulates, and Principles, meant that several high-level attributes or facets could be used to describe any entity (in his work, the entities were books). Ranganathan’s five initial facets were “Personality”, “Matter”, “Energy”, “Space” and “Time” but it is apparent that now, these terms, although appropriate for Ranganathan and librarians of the time, are not useful for all users of books. Automated methods for identifying facets are now being investigated (Ben-Yitzhak et al., 2008).

When comparing the results of this study with the faceted navigation structure utilised by Amazon for the "Music" section of its website, it can be seen that there are some similarities and significant differences.

The number of similarities demonstrates that Repeated single-criterion sorts are a potential method for eliciting these types of navigation structure. The search filters that cannot be mapped onto a specific construct can be explained by the choice of entities used in this study. The card sorting tool only provided respondents with the mp3 of the song itself. No information was given regarding the artist, edition, or format, so it is unlikely that these could have been used as criteria. Therefore, choice of the representation for a song needs further consideration and a mix of media may be more appropriate e.g. having the album cover representing the song instead of a number or a screenshot of the product description webpage (which could then also incorporate price and delivery options).

More noteworthy is that there are a number of constructs elicited during this study that are not part of the search filters on the Amazon website, specifically "Gender of Artist", "Speed of song", "Solo or group", "Main instrument", "Audience", "Emotion" and "Nationality of artist". These are all potential methods of guiding a user through a set of search results and apart from "Emotion" are all objective characteristics of the entity. This is similar to the findings of Cassidy (Cassidy et al., 2013) who used "All-in-one" sorting to determine how children categorise games. When compared with existing categories in the Google Play Store, they found that "children used categorization criteria much more aligned to the goals of the game rather than more abstract categories currently found in mobile phone application stores" (Cassidy et al., 2013).

6.5 Implications for the UCD process

"All-in-one" card sorting is already used in the UCD process to determine the IA of websites. However, this is commonly used to determine single level hierarchies where an entity fits into one specific top-level category e.g. Books, Music, Games, Films.

Repeated single-criterion sorting is a complimentary method for eliciting faceted navigation structures for when a user has chosen a top-level category, such as Books, and is now looking for a particular item using relevant criteria such as Author, Publication Date, Genre etc. From this study it is clear that this method can elicit traditional criteria which are akin to those originally proposed by Ranganathan but can also provide additional, user-centered dimensions such as those seen in user generated tagging systems.

A combination of the two sorting variations would provide a methodology for developing IA's that avoid the limitations of hierarchical structures where products/pages may fit into multiple sub-categories. Closed sorting could also be used with both sorting variants to evaluate how well the category and criteria labels work.

Repeated single-criterion sorting could also be used in the initial stages of user experience design to analyse the target users' perceptions of web sites in the particular field. For example, if a developer was creating a site for a theatre company, the tool could be populated with images of the homepages of other theatre company websites and a study into users' perceptions of those sites could be undertaken. The results from this study could then be fed directly into the development process, with attention then being directed at the attributes of the pages that were elicited.

6.6 Challenges and Limitations

The main challenge with repeated single-criterion sorting is the same as with any other user-centered activity; recruiting respondents. For this study, a pool of willing participants was fortunately available, but this in itself causes potential issues with bias and representativeness. It should also be noted that there is limited published evidence on how many respondents are needed for this type of sorting to create effective IA's and this is where future work is needed i.e. to take the results from a study such as this, build a website with faceted navigation and then evaluate it with users.

7. CONCLUSIONS

The results of this study have demonstrated that card sorts are an effective method for investigating people's perceptions of music. Repeated single-criterion music sorts have elicited criteria which have previously eluded music psychologists and digital music service designers, while providing explanations to inconsistent musical genre classifications seen in previous studies.

Current support for digital media organisation and discovery relies heavily on genre across a range of different product types e.g. Music, Books and Films. Although genre was used by the majority of respondents, there was little agreement between respondents in which songs should appear in the same genre. This study has shown that perception of music is highly subjective, and genre, although considered to be objective by music retailers, is no longer adequate. The results indicate that as the volume, and variety of music increases, categorising music becomes more difficult. Repeated single-criterion sorting provides a method to support the systematic elicitation of additional objective and

subjective features for use in the design of digital music services. This variation of card sorting has previously been reliant on a number of time-consuming processes (recruiting respondents, conducting synchronous sessions etc.) and although low cost and high yield, has not been frequently used as a standard part of design and development methodologies. The key contribution of this work therefore is a demonstration that this card sorting variation and tool described can support developers and designers at various stages in the development life cycle to determine IA's for websites and applications that go beyond traditional objective features. This study has also shown how this method could be used as part of the UCD process in parallel with the more common "All-in-one" version to create faceted navigation structures.

9. ACKNOWLEDGEMENTS

Thanks to Dr Gordon Rugg for his supervision during the data collection stages of this study.

10. REFERENCES

Aucouturier, J.J. & Pachet, F. (2003). Representing musical genre: A state of the art. *Journal of New Music Research*, 32,1: 83-93.

Ben-Yitzhak, O., Golbandi, N., Har'El, N., Lempel, R., Neumann, A., Ofek-Koifman, S., Sheinwald, D., Shekita, E., Sznajder, B. & Yogev, S. (2008). Beyond basic faceted search. In *Proceedings of the international conference on Web search and web data mining*, 33- 44.

Bevan, N. (2003). *UsabilityNet methods for user centred design. Human-Computer Interaction: theory and Practice 1*. APA.

Cassidy, B., Saurabh Antani, S. & Read, J. (2013). Using an open card sort with children to categorize games in a mobile phone application store. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, NY, USA, 2287-2290.

Courage, C. & Baxter, K. (2005). *Understanding your users: a practical guide to user requirements: methods, tools, and techniques*. San Francisco, CA: Morgan Kaufmann Publishers.

de Quincey, E. (2010). *Software support for comparison of media across domains*. Ph.D. Dissertation, Keele University, UK.

Deibel, K., Anderson, R. & Anderson, R. (2005). Using edit distance to analyze card sorts. *Expert Systems*, 22,3: 129-138.

DeNora, T. (2000). *Music in everyday life*. Cambridge: Cambridge University Press.

Frederickson-Mele, K. (1997). Usability Testing an Intranet Prototype Shell: A Case Study. In *Proceedings of SIGCHI 1997*.

Gammack, J.G. (1987). Different Techniques and Different Aspects of Declarative Knowledge. In *Knowledqe Acquisition for Expert Systems*. New York: Plenum Press, 137-163.

Gerrard, S. & Dickinson, J. (2005). Women's working wardrobes: a study using card sorts. *Expert Systems*, 22,3: 108-114.

Golder, S. & Huberman, B. (2005). *The structure of collaborative tagging systems*. HP Labs Technical Report.

Greasley, A. & Lamont, A. (2006). Music preference in adulthood: Why do we like the music we do. In *Proceedings of the 9th International Conference on Music Perception and Cognition*.

Hudson, W. (2014). Card sorting for navigation design. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '14). ACM, New York, NY, USA, 1007-1008.

Joyce, J. (2006). Pandora and the Music Genome Project. *Scientific Computing,* 23,10: 40-1.

Kipp, M. (2007). @toread and cool: Tagging for time, task and emotion. In *Proceedings Information Architecture Summit 2007*.

Kipp, M. & Campbell, D.G. (2006). Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. In *Proceedings of the American Society for Information Science and Technology*, 43,1.

Lippens, S., Martens, J-P., & De Mulder, T. (2004). A comparison of human and automatic musical genre classification. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004* (ICASSP'04), 4.

Maiden, N. (2009). Card Sorts to Acquire Requirements. *IEEE Software*, 26,3: 85-86.

Martine, G. & Rugg, G. (2005). That site looks 88.46% familiar: quantifying similarity of Web page design. *Expert Systems,* 22,3: 115-120.

McKay, C. (2010). *Automatic music classification with jMIR*. Ph.D. Dissertation, McGill University, Canada.

Nakhimovsky, Y., Schusteritsch, R. & Rodden, K. (2006). Scaling the card sort method to over 500 items: restructuring the Google AdWords help center. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '06). ACM, New York, NY, USA, 183-188.

Nielsen, J. (1992). The usability engineering life cycle. *Computer*, 25,3: 12-22.

Nielsen, J. & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People* (CHI '90). ACM, New York, NY, USA, 249-256.

North, A.C., Hargreaves, D.J. & O'Neill, S.A. (2000). The importance of music to adolescents. *British Journal of Educational Psychology*, 70,2: 255-272.

O'Neill, D. (2013). Retrieved September 16, 2016, from http://id3.org/

Pachet, F. & Cazaly. D. (2000). A taxonomy of musical genres. In *Proceedings of Content- Based Multimedia Information Access (RIAO)*, 1238-1245.

Ranganathan, S.R. (1962). *Elements of library classification*. Bombay: Asia Publishing House.

Rugg, G. Corbridge, C., Major, N. & Burton, A. (1992). A comparison of sorting techniques in knowledge acquisition. *Knowledge Acquisition,* 4,3: 279-291.

Rugg, G. & McGeorge, P. (1997). The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems*, 14,2: 80-93.

Sanders, K., Fincher, S. Bouvier, D., Lewandowski, G., Morrison, B., Murphy, L., Petre, M. et al. (2005). A multi-institutional, multinational study of programming concepts using card sort data. *Expert Systems,* 22,3: 121-128.

Scheirer, E. (1998). Tempo and beat analysis of acoustic musical signals. *The Journal of the Acoustical Society of America*, 103,1: 588-601.

Sloboda, J., O'Neill, S.A. & Ivaldi, A. (2001). Functions of music in everyday life: An exploratory using the experience sampling method. *Musicae Scientiae*, 5,1: 9- 32.

Spencer, D. & Warfel, T. (2009). Card Sorting: A Definitive Guide. Retrieved January 18, 2016, from http://www.boxesandarrows.com/view/card_sorting_a_definitive_guide.

Tullis, T. (1985). Designing a menu-based interface to an operating system. *ACM SIGCHI Bulletin*, 16,4: 79-84.

Tullis, T. (2003). Using Card-sorting Techniques to Organize Your Intranet. *Intranet Journal of Strategy and Management*, March 2003.

Tullis, T. & Wood, L. (2004). How Many Users Are Enough for a Card-Sorting Study? In *Proceedings of the UPA 2004 Conference*, Minneapolis, MN, USA. Jun-2004.

Upchurch, L., Rugg, G. & Kitchenham, B. (2001). Using card sorts to elicit web page quality attributes. *IEEE Software,* 18,4: 84-9.

Usability.gov. (N.D.). *User-Centered Design Process Map*. Retrieved September 16, 2016, from http://www.usability.gov/how-to-and-tools/resources/ucd-map.html.

Wharton, C., Rieman, J., Lewis, C. & Polson, P. (1994). The cognitive walkthrough method: A practitioner's guide. In *Usability Inspection Methods*, J. Nielsen and R. L. Mack, eds. New York, NY: John Wiley & Sons, 105-140

Digital Wellbeing for All: Expanding Inclusivity to Embrace Diversity in Socio-Emotional Status

Reem S. Al-Mansoori
College of Science and Engineering
Hamad Bin Khalifa University
Doha, Qatar
reem.almansoori.qtr@gmail.com

Mohammad Naiseh
School of Electronics and Computer Science
University of Southampton
Southampton, UK
m.naiseh@soton.ac.uk

Dena Al-Thani
College of Science and Engineering
Hamad Bin Khalifa University
Doha, Qatar
dalthani@hbku.edu.qa

Raian Ali
College of Science and Engineering
Hamad Bin Khalifa University
Doha, Qatar
raali2@hbku.edu.qa

The ubiquity of information and communication technology contributed positively in enhancing lives, mainly in increasing productivity and economic growth, while their impact on life satisfaction and wellbeing has been a hidden cost. Digital media shall empower users to maximise their digital wellbeing, i.e. healthy and regulated relationship with technology. Similar to usability, people differ in their needs to achieve and maintain their digital wellbeing. A technology design shall be inclusive in how it helps users to increase their digital wellbeing and reduce possible harm. Typical inclusivity dimensions in Human-Computer Interaction research include gender, race, physical and cognitive abilities, with the aim of making the product usable by the wider possible user set. However, another range of inclusivity dimensions becomes prominent and that is the diversity in users' socio-emotional characteristics such as susceptibility to online pressure (technical and social), resilience and others. Such characteristics can be traits, e.g. introversion, or temporal status, e.g. being in a low mood. In this position paper, we are proposing digital wellbeing as a target for an inclusive design where technology designers need to anticipate and reduce the negative impact of their products and services on the wellbeing of users through considering their diverse socio-emotional status.

Keywords: Human-Computer Interaction, inclusivity, digital wellbeing.

1. INTRODUCTION

As technology permeates every aspect of our daily lives, having technology that is useful, inclusive and safe is essential. In Human-Computer Interaction (HCI), the relationship between humans and digital products has evolved from a desire for instant and intuitive screen interaction to one where such ubiquity and speed started to impact humans' wellbeing (Verbeek, 2005).

As information and communication technology (ICT) outputs occur at a rapid pace, it is difficult to predict innovation outcomes in this area compared to other areas of science. Software may be released to the users and 'go viral', causing tremendous effects on individuals' activities and societies' structures that have not been anticipated by their designers (Jirotka et al., 2017). In the case of social media, for instance, a number of studies showed the negative impacts on users' mental health correlated with the use of social media problematic usage (Wright et al., 2018, Marino et al., 2018). To support more balanced relationship with technology and help users' mental health and wellbeing, companies started to launch dedicated toolkits such as Google's Digital Wellbeing (https://wellbeing.google/). These toolkits cover basic wellbeing properties, e.g. minimising distraction, pausing and night mode for quality sleep. For this reason, and to minimise possible mental and cognitive costs triggered by certain interactive design features, digital platforms, services and devices need to anticipate the impact on users' wellbeing and increase digital wellness to the wider possible set of users and context of use.

In this position paper, we discuss the need for digital wellbeing services and toolkits to be inclusive and sensitive to the diversity in the socio-emotional characteristics of users, whether traits or temporal and contextual. Our definition of social-emotional status is based on the work of Halberstadt et al., 2011 and the attributes contributing to it proposed in (Parhomenko, 2014). In this context, we identify socio-emotional status as the degree to which an individual is able to practice conscious behaviour regulation, build harmonious relationships and be engaged in social activities. Being socio-emotional competent represents possessing the abilities to be self-aware, empathetic, motivated (to be involved even when facing difficulties or failures), self-regulated (having conscious self-control over emotions) with social knowledge and competencies

to build and maintain healthy relationships (Parhomenko, 2014).

We provide a background on digital wellbeing and inclusive design, in addition to the paper's motivation in Section 2. Taking social media as an exemplar, we explain how popular design features lack sensitivity to users' diversity in their socio-emotional status in Section 3. We discuss our proposal and its significance in Section 4 and we present challenges and future work directions in Section 5.

2. BACKGROUND AND MOTIVATION

Digital wellbeing is defined, by Widdicks et al., 2017, as "a positive feeling associated with the use of technology, striven by maintaining a balance between our 'real' and 'online' lives". Although this notion is relatively new, a considerable body of research had provided a foundation for it. This includes works on positive technology (Riva et al., 2012), positive design (Desmet and Pohlmeyer, 2013), positive computing (Calvo and Peters, 2014), experience design (Hassenzahl, 2010) and the work on Motivation Engagement and Thriving User Experience (METUX) in (Peters et al., 2018). The design frameworks resulted from these efforts share the goal of satisfying a number of users' psychological needs, such as autonomy, self-awareness, gratitude and ultimately enhance their wellbeing.

Inclusive design is a "project that sets out to include significant sectors of society that are all too frequently ignored or overlooked" (Coleman et al., 2003). Morris (2003) and Coleman et al. (2003) emphasised that inclusive design should address the needs of the widest possible audience, irrespective of age or ability. Rossetti, on the other hand, defined universal design (a term equivalent to inclusive design and used mostly in the U.S and Japan) as "the good design", and provided seven guiding principles (Rossetti, 2006). Heylighen and Bianchin (2013) explored this perspective further and discussed how the characteristics of an inclusive design are similar to those of a "good design".

Inclusivity has been often seen as the design ability to cater to the common forms of disabilities, such as visual and hearing impairment and physical disabilities. Others approached inclusivity from other perspectives, such as "diversity dimensions" (Himmelsbach et al., 2019). Age, gender, ethnicity, physical and mental abilities were extensively examined in HCI literature responding to calls for diversity-sensitive research (Himmelsbach et al., 2019).

Inclusivity definitions and frameworks varied from being general (i.e. an enabler for the diverse users rather than an excluder (Gyi et al., 2000)), to having the focus on physical, sensory, cognitive abilities, or all (Persad et al., 2007; Keates, and Clarkson, 2003; Abascal and Azevedo, 2007), to being socially inclusive through addressing social differences, such as inequalities (classes) which affect the social and structural contexts of how people experience technology (Himmelsbach et al., 2019; Erete et al., 2018). The consideration of the socio-emotional status of the user is yet to be seen as an inclusivity factor.

In our proposal for an inclusive design with the target of safeguarding users' digital wellbeing, we suggest that the design shall provide measures to avoid users' negative feelings of exclusion, isolation, anxiety and fears. For the design to be inclusive, it shall consider the diversity in the socio-emotional status. For example, in applying this vision to social media, a wellbeing-aware design shall be accountable to those prone to be overly influenced by peer pressure and lack self-esteem. Such a design can tailor content and offer services to configure and customize interfaces to avoid causing anxiety and sadness. Possible interventions include diversifying the content, reminding users of the nature of interactions in cyberspace, providing users with tools for empowered refusal should they wish to opt-out, issuing direct or subliminal messages to promote healthy usage, and others. Design frameworks for inclusive digital wellbeing should aid designers in knowing how the diversity in the socio-emotional status inter-relates with design options and the context of use, and how to mitigate the negative impact.

3. DESIGN FEATURES AND WELLBEING

The challenges surrounding inclusive design remain in the fact that designers need generalization to make it a universal design, and at the same time, they need specifications to account for humans' individuality (Luck, 2018). Designing for an average user, which is an artificial construct, is necessarily exclusive as average can be claimed as 'statistical myth' (Rose, 2016). Algorithmic-based designs, for instance, are biased as they promote content based on techniques such as collaborative filtering. This content customising process can promote content that is popular and make other content less visible, hence, excluding its audience. Different people possess diverse characteristics and variable capabilities, therefor, designers need to anticipate users' lived experiences (Luck, 2018). If we design with an average user in mind, many users may need to adapt to the interface provided to them (Abascal and Azevedo, 2007). This will result in a number of

users not being able to do so and facing feeling of exclusion, which will affect their wellbeing.

The notion of the average user becomes even more questionable when addressing the socio-emotional characteristics of users. Emotional wellbeing is part of digital wellbeing and has been identified as a fundamental indicator of the quality of life as per the World Health Organization (Coyle et al., 2012). In social media designs, taken as an exemplar in this paper, features such as emoticons (in Facebook) are among the reasons why some individuals may suffer from negative emotions such as jealousy, which, as a consequence, lead to reduced wellbeing (Hudson et al., 2015). When these features are used by individuals with low self-esteem and high susceptibility to social influence, negative emotions and psychological states are more likely to arise (Altuwairiqi et al., 2019a).

In certain situations, like being in a remote geographic area or in lockdown, e.g. those related to the current COVID-19 pandemic, social media become highly needed and extensively consumed for essential wellbeing needs such as relatedness and feeling secure and informed. It can help some groups of users to maintain connectivity with others and maintain resilience in days of isolation. Beyond connectedness, people seek acceptance. It is a human nature to seek social acceptance, which is one of the determinants of wellbeing (DeWall and Bushman, 2011). Online features such as Likes can help enhance or deteriorate individuals' wellbeing as users may overly rely on them as indicators of the extent to whether they are socially accepted. The opposite of social acceptance, rejection, on the other hand, can increase anxiety, anger, depression and jealousy (Leary, 2010), and diminish an individual's self-esteem (Williams et al., 2000). Moreover, rejection is said to increase selfish and aggressive behaviour, leading to mass violence (Leary et al., 2003). Conversely, engaging these individuals in social activities and offering a sense of acceptance showed to reduce their aggression (DeWall et al., 2010).

Interacting with others on social media provides a context where individuals are subject to forms of social evaluation threats, in which rejection is one of them (DeWall and Bushman, 2011). In the physical world, being exposed to negative external judgements by others can lead to releasing stress hormone cortisol (Dickerson and Kemeny, 2004) and stimulating the production of proinflammatory cytokines (Dickerson et al., 2009), which are negative biological responses affecting humans' wellbeing. Studies showed that individual factors contribute to heightened and diminished responses to social rejection, hence differ from one person to another (DeWall and Bushman, 2011). Rejection and social disapprovals showed high neural correlations among people with high rejection sensitivity (Burklund et al., 2007) and low self-esteem (Onoda et al., 2010). Although cyber behaviour is not necessarily analogous to that in the physical world and we still require more research to understand its nuances, studies are showing similarity, e.g. in the fear of ostracism (Vorderer and Schneider, 2017) and the neural impact of certain social media features, such as Likes, on users and their wellbeing (Sherman et al., 2016).

4. INCLUSIVE DESIGN FOR WELLBEING

Some might argue that individuals ought to manage their own traits, weaknesses, insecurities and anxieties through self-regulation. However, as the design can make the user experience more compelling (and sometimes addictive), it can also consider reducing the triggers that underpin compulsive behaviours (Churchill, 2020), such as fear of missing out (FoMO). FoMO is a type of addictive online behaviour and represents an over-dependence on social media due to the fear of not seeing or reacting to online activities on time (Alutaybi et al., 2019). Digital addiction can lead to anxiety, interrupted sleep, lack of concentration and emotional control (Cham et al., 2019, Altuwairiqi et al., 2019b). As such, FoMO is an example of how the design, in a particular socio-emotional status and context of use, triggers negative feelings such as fear. At the same time, an inclusive design can provide design features such as an advanced version of the autoreply, status and filtering (Alutaybi et al., 2019) to empower those who may be overly susceptible to social pressure and lack impulse control. Digital wellbeing can be supported by software mediated countermeasures which target behaviours directly and indirectly. This includes using limits settings (Löchtefeld et al. 2013), or virtual reality (Park et al., 2016), where the former focuses on usage restriction rules and time, and the latter on pairing the addictive behaviour with irritating noise (or other features) to associate it with negative feelings, hence reduce excessive gameplay.

We hypothesise that for a design to be inclusively *wellness-sensitive*, it has to address three essential inclusivity parameters, which are: physical, cognitive and socio-emotional status (Figure 1). We advocate that digital media design should be inclusive not only in terms of being accessible and usable to all and in a wide variety of contexts but also in accommodating the diversity in the socio-emotional status of users as a requirement for wellbeing.

The base of this pyramid represents the first encounter between humans and computers through accessibility, where access to digital platforms and services is guaranteed for users and caters for their diverse physical and cognitive abilities (i.e., visual impairment and hearing loss). Then comes usability, where people can use a design product (or service) to achieve specified goals with effectiveness, efficiency and satisfaction in a

specified context of use (ISO, 2009). On the top, and closer to the users' socio-emotional status, comes the user experience which considers the satisfaction of users' pragmatic and hedonic goals (Petrie and Bevan, 2009). Optimising users' performance and users' satisfaction by achieving goals (ISO, 2009) are considered parts of a successful user experience that aims to enhance users' wellbeing. Our suggestion is to go a step further to consider digital wellbeing and be inclusive in accommodating the variation in users' socio-emotional status (permanent and situational). In this way, not only we design to enhance users' experience, but also we aim to protect their wellbeing and mental health when interacting with technology, e.g. by helping them balance between 'real' and 'online' lives, hence, safeguarding their wellbeing (Widdicks et al., 2017).

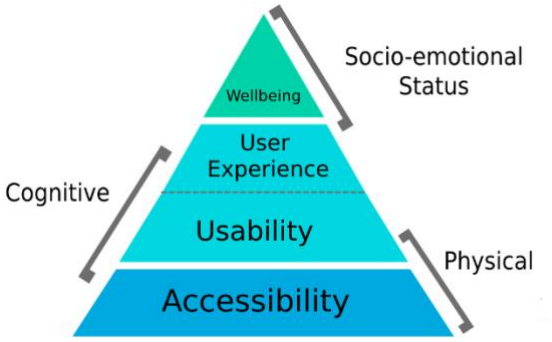


Figure 5: Inclusivity pyramid with wellbeing as a target

5. CHALLENGES AND FUTURE RESEARCH OPPORTUNITIES

Digital inclusion is not a straightforward topic. Perlgut, 2011, predicted that digital exclusion would be the major social justice challenge that would surpass all other social and economic determinants. The current work sheds light on inclusivity in its broader sense (beyond race, gender, physical and cognitive disabilities). This work suggests modifying inclusivity indicators to consider people's diverse socio-emotional status as an attempt to safeguard their wellbeing. To the best of our knowledge, this work is the first in proposing diversity in socio-emotional status as another parameter to inclusivity.

While inclusivity (considering socio-emotional status) with the aim of increasing digital wellbeing is a desirable design goal, a number of challenges are to address. The decision on the requirements and the design to achieve this goal requires a lengthy process involving users of different socio-emotional abilities that could be simulated in different context of usage. Drawing conclusions from the data collected will be also difficult in terms of generalisability, similar to research studies. Best practices and guidelines are yet to formulate and evidence of their value is yet to generate. The literature on cyberpsychology is still emerging and studies are mainly focused on the user and their behaviour on digital media rather than the design and its nuances and features (and mostly platform-dependent). Each cyberspace exhibits its own instantiation making it hard to assume results are replicable (e.g. the Like's meaning depends on the spirit of the platform and the expectations amongst peers and their norms). In addition, we need measurement tools and ways to establish liability when it comes to assessing the role of design and use in altering the socio-emotion status, and vice versa. Designers' assumptions and decisions can be biased to what they would consider as an average user or a range of possible user models and personas. This can be still a challenge especially when addressing the relationship between socio-emotional status and design space variability. As the design and testing process may entail provoking negative emotions in users, ethical concerns might arise.

Designing and testing for inclusivity (considering the variation in users' socio-emotional status) and their inter-relation with the design and its use is a complex question to answer. The main challenge entails the diversity in users' status and the variability in the design, the context of use and their combinations. Overcoming this challenge can benefit from the novel data gathering and adaptive interfaces where artificial intelligence is used to recognise users' emotional status, put smart barriers and protect the needs of users (Churchill, 2020). Such a process can be applied, iteratively, to help designers to optimize their designs through the lived experience and the data captured to respond to the emerging wellbeing requirements.

3. REFERENCES

Abascal, J, and Azevedo, L. (2007). "Fundamentals of Inclusive HCI Design". In C. Stephanidis (Ed.): Universal Access in HCI, Part I, HCII 2007, LNCS 4554, pp. 3–9, 2007. © Springer-Verlag Berlin Heidelberg 2007

Altuwairiqi, M.; Jiang, N.; Ali, R. Problematic Attachment to Social Media: Five Behavioural Archetypes. International Journal Environmental Research And Public Health (2019a), 16, 2136.

Altuwairiqi, M., Arden-Close, E., Bolat, E., Vuillier, L. and Ali, R., (2019b). When People are Problematically Attached to Social Media: How Would the Design Matter?. In Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC 2019), Bari, Italy

Alutaybi, A.; McAlaney, J.; Arden-Close, E.; Stefanidis, A.; Phalp, K.; Ali, R. Fear of Missing Out (FoMO) as Really Lived: Five Classifications and one Ecology. In Proceedings of the 2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC), Beijing, China, 28–30 October 2019; Institute of

Electrical and Electronics Engineers (IEEE): Miami, FL, USA, 2019; pp. 1–6

Burklund, L.J., Eisenberger, N.I., & Lieberman, M.D. (2007). The face of rejection: Rejection sensitivity moderates dorsal anterior cingulate activity to disapproving facial expressions. Social Neuroscience, 2, 238–253.

Calvo, R., Peters, D. (2014) Positive Computing: Technology for Wellbeing and Human Potential. The MIT Press. Cambridge, Massachusetts

Cham, S.; Algashami, A.; Aldhayan, M.; McAlaney, J.; Phalp, K.; Almourad, M.B.; Ali, R. Digital Addiction: Negative Life Experiences and Potential for Technology-Assisted Solutions. Adv. Intell. Syst. Comput. 2019, 16, 921–931.

Churchill, E. (2020). Designing for digital wellbeing. ACM Digital Library. https://interactions.acm.org/archive/view/january-february-2020/designing-for-digital-well-being

Coleman, R., Lebbon, C., Clarkson, P., and Keates, S. (2003). From margins to mainstream: why inclusive design is better design. In P. Clarkson (Ed.), Inlusive design: Design for the whole population (pp. 1-25). London: Springer

Coyle, D., Linehan, C., Tang, K., Lindley, S. (2012). Interaction design and emotional wellbeing. CHI 2012, May 5–10, 2012, Austin, TX

Desmet, P., and Pohlmeyer, A. (2013). Positive Design An Introduction to Design for Subjective Well-Being. International Journal of Design. 7, 5-19

DeWall, C.N., Bushman, B.J. (2011). Social acceptance and rejection: the sweet and the bitter. Current Directions in Psychological Science. 20(4), 256-260

DeWall, C.N., Twenge, J.M., Bushman, B.J., Im, C., & Williams, K.D. (2010). Acceptance by one differs from acceptance by none: Applying social impact theory to the rejection–aggression link. Social Psychological and Personality Science, 1, 168–174

DeWall, C. N., & Bushman, B. J. (2011). Social acceptance and rejection: The sweet and the bitter. Current Directions in Psychological Science, 20(4), 256-260.

Dickerson, S.S., & Kemeny, M.E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. Psychological Bulletin, 130, 355–391

Dickerson, S.S., Gable, S.L., Irwin, M.R., Aziz, N., & Kemeny, M.E. (2009). Social-evaluative threat and proinflammatory cytokine regulation: An experimental laboratory investigation. Psychological Science, 20, 1237–1244.

Erete, S, Israni, A, Dillahunt, T (2018). “An Intersectional Approach to Designing in the Margins”.

Gyi, D.E., Porter, J., Case, K. (2000). Design practice and designing for all. In Human Factors and Ergonomics Society Annual Meeting Proceeding. San Diego, CA, 913-916

Halberstadt, A.G., Denham, S.A., & Dunsmore, J.C. (2001). Affective social competence, Social Development; 2001, 10, 79 - 119.

Hassenzahl, M (2010) Experience design: Technologu for all the right reasons. Morgan and Claypool Publisher.

Heylighen, A., Bianchin, M. (2013). How does inclusive design relate to good design? Designing as a deliberative enterprise, Design Studies. 34 (1), 93-110

Himmelsbach, J., Schwarz, S., Gerdenitsch, C., Wais-Zechmann, B. & Tscheligi, M. (2019). Do We Care About Diversity in Human Computer Interaction. A Comprehensive Content Analysis on Diversity Dimensions in Research. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems – CHI ’19. https://doi.org/https://doi.org/10.1145/3290605.3300720

Hudson, M. B., Nicolas, S. C., Howser, M. E., Lipsett, K. E., Robinson, I. W., Pope, L. J., Friedman, D. R., et al. (2015). Examining how gender and emoticons influence Facebook jealousy. *Cyberpsychol. Behav. Soc. Network.* 18, 87–92. doi: 10.1089/cyber.2014.0129

ISO FDIS 9241-210 (2009). Human-centred design process for interactive systems. ISO.

Jirotka, M, Grimpe, B, Stahl, B., Eden, G., Hartswood, M. (2017). Responsible Research and Innovation in the Digital Age. Communications of the ACM. 60, 62-68

Keates, S., Clarkson, J. (2003) Countering design exclusion—an introduction to inclusive design. Springer, Heidelberg

Widdicks, K., Bates, O., Hazas, M., and Friday, A. (2017). Designing for Digital Wellbeing and its Challenges in Everyday Life. In Proceedings of ACM Conference (Conference’17). ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Leary, M.R. (2010). Affiliation, acceptance, and belonging. In S.T. Fiske, D.T. Gilbert & G. Lindzey (Eds.), Handbook of Social Psychology (5th Ed., Vol. 2, pp. 864–897). New York, NY: Wiley

Leary, M.R., Kowalski, R.M., Smith, L., & Phillips, S. (2003). Teasing, rejection, and violence: Case studies of the school shootings. Aggressive Behavior, 29, 202–214.

Löchtefeld, M., Böhmer, M. and Ganev, L., (2013). AppDetox: helping users with mobile app addiction. *Proceedings of the 12th International Conference on Mobile and Ubiquitous*

Multimedia, Luleå, Sweden. Association for Computing Machinery. Article 43.

Luck, R. (2018). Inclusive design and making in practice: Bringing bodily experience into closer contact with making. Design Studies (54). 96-119

Marino, C., Gini, G., Vieno, A., Spada, MM (2018). The associations between problematic Facebook use, psychological distress and well-being among adolescents and young adults: A systematic review and meta-analysis. Journal of Affective Disorder. 15 (226), 274-281. doi: 10.1016/j.jad.2017.10.007. Epub 2017 Oct 3. PMID: 29024900.

Morris, L. (2003). Preface. Inclusive design: design for the whole population. In J. Clarkson, R. Coleman, S. Keates, and C. Lebbon (Eds.). London: Springer-Verlag

Onoda, K., Okamoto, Y., Nakashima, K., Nittono, H., Yoshimura, S., Yamawaki, S., et al. (2010). Does low self–esteem enhance social pain? The relationship between trait self–esteem and anterior cingulate cortex activation induced by ostracism. Social Cognitive & Affective Neuroscience, 5, 385–391

Parhomenko, K. (2014). Diagnostic Methods of Socio – Emotional Competence in Children. Third annual international conference (Early childhood care and education). Procedia - Social and Behavioral Sciences 146 329 – 333

Park, S. Y., Kim, S. M., Roh, S., Soh, M.-A., Lee, S. H., Kim, H., Lee, Y. S. and Han, D. H., (2016). The effects of a virtual reality treatment program for online gaming addiction. *Computer Methods and Programs in Biomedicine*, 129, 99-108.

Perlgut, D. (2011). Digital Inclusion in the Broadband World: Challenges for Australia.? presented at the Communications Policy and Research Forum, Sydney, November 2011.

Persad, U., Langdon, P., Clarkson, J. (2007). Characterising user capabilities to support inclusive design evaluation. Universal Access in the Information Society. 6, 119-135.

Peters D, Calvo RA and Ryan RM (2018). Designing for Motivation, Engagement and Wellbeing in Digital Experience. Front. Psychol. 9:797. doi: 10.3389/fpsyg.2018.00797

Riva, G., Banos, R., Botella, C., Wiederhold, B., Gaggiolo, A. (2012). Positive Technology: Using Interactive Technology to Promote Positive Functioning. Cyberpsychology, Behavior and Social Networking. 15, 69-77.

Rose, T. (2016). The End of Average: How we Succeed in a World that Values Sameness. San Francisco, CA, HarperOne

Rossetti, R. (2006). The seven principles of universal design. Action Magazine. https://www.udll.com/media-room/articles/the-seven-principles-of-universal-design/

Sherman, L. E., Payton, A. A., Hernandez, L. M., Greenfield, P. M., & Dapretto, M. (2016). The power of the like in adolescence: Effects of peer influence on neural and behavioral responses to social media. Psychological science, 27(7), 1027-1035.

Verbeek, P. P. (2005). What things do: Philosophical reflections on technology, agency, and design. (R. P. Crease, Trans.). University Park, PA: Penn State Press.

Vorderer, P., & Schneider, F. M. (2017). Social media and ostracism. Ostracism, exclusion, and rejection, 240-257.

Williams, K.D., Cheung, C.K.T., & Choi, W. (2000). Cyberostracism: Effects of being ignored over the Internet. Journal of Personality and Social Psychology, 79, 748–762

Wright E.J., White, K.M., Obst, P.L (2018). Facebook False Self-Presentation Behaviors and Negative Mental Health. Cyberpsychology Behaviour and Social Networking. 21(1), 40-49. doi: 10.1089/cyber.2016.0647.

Design Discovery Practices: Engaging professional design communities with Ketso

Kevin Rigotti
UCLan
Preston
KDRigotti@uclan.ac.uk

Peggy Gregory
UCLan
Preston
AJGregory@uclan.ac.uk

Dan Fitton
UCLan
Preston
DBFitton@uclan.ac.uk

Decisions and assumptions made during design sessions, when teams are formulating their design objectives and their understanding of the problem they intend to solve, can be essential to the outcome as they fundamentally shape and direct the design of the product or service that is delivered. Current practice in these crucial design discovery activities is under-explored in the academic literature. To address that shortfall, and answer the research question of how UX practitioners approach and perform discovery, we used the Ketso workshop format to explore the design discovery process and its challenges with 12 user researchers and designers from a university and a large retail organisation. Our thematic analysis of the workshop outputs showed that practitioners valued an empirical data-led approach, where they could have confidence in the coverage and validity of the data, and achieve a shared understanding of the user research findings across the organisation. Key challenges included the mindset of stakeholders, with whom practitioners wanted deeper engagement, and constraints on time which may require HCI research to develop practical solutions.

Workshops, Engagement, Communities of Practice, Thematic analysis, Design discovery

1. INTRODUCTION

Choices made during design sessions may be based on an incomplete, and possibly mistaken, understanding of the problem, but still have the potential to fundamentally shape and direct the design in ways that will be apparent in the product or service that is delivered to customers. Some choices are conscious decisions, others may be unchallenged assumptions. In order to have confidence in our delivery processes, we first need to have confidence in how we decided what to build.

Teams formulate their design objectives, and their understanding of the problem they intend to solve, by establishing that there is a need to be met, that they know how to build it, that potential users will want it, and that stakeholders will support it. The data required for this comes from user research and business analysis confirming the viability of a product. These activities are collectively termed ‘design discovery' (Brown, 2017) or ‘product discovery' (Cagan, 2018) or most commonly in the UX practitioner community simply ‘discovery' (Government Digital Service, 2019). Use of the term ‘discovery' in academic literature is more limited, typically it is used to refer to business models, as in ‘discovery driven' (McGrath, 2010), or when discussing Lean start-up approaches (Shepherd and Gruber, 2020).

Academic literature describing design discovery practices in the UK software industry, and the rationale for their choice, is sparse and not specific to a geographic community of practice. The available papers take their focus from particular segments of the client community rather than the practitioner community, such as addressing particular issues for UX with children (Sim et al., 2017), or are concerned more with the integration of design with development than with design itself, for example examining the relationship with Agile as described by Salah et al. (2014) or Gregory et al. (2015).

To address that shortfall, the research question for this study is

> How do UX practitioners approach and perform discovery?

In order to address that question, workshops were held in summer 2019 and early 2020 to gather information on current design discovery practice by exploring the objectives that practitioners aim to satisfy. The scope of the workshops was specifically the design discovery stage of development, covering the planning and conduct of user research and the initial presentation and shared understanding of the research findings by the project team as a whole. While recognising that discovery implies a learning curve, and design choices may be made at different points of that curve, no assumptions were made about the distribution of discovery activity over time, and it was not assumed to occupy a distinct phase of development.

This paper contributes an initial thematic analysis of the key aspects of design discovery used by practitioners, the best practice that practitioners aspire to, and the organisational challenges that they identified.

2. RELATED WORK

The desire for rapid iterative deployment of software and other digital products has placed challenging demands on the user research and design activities that precede and support the product development. Various attempts have been made to integrate User Experience (UX) design with continuous development in ways that achieve greater agility. Some have taken a dual track approach where design activity runs in parallel but synchronises at key points (Cagan, 2018, Trieflinger et al., 2021), or periodically inserts a short design sprint to answer key design questions (Knapp et al., 2016). Other approaches, drawing on Lean manufacturing ideas (Gothelf and Seiden, 2016), have posed incremental outcome hypotheses to be tested by each development sprint.

As the available methods differ so significantly, it is important to understand which are being used and how they are being applied in practice. An international survey of understandings of User Experience (UX) design by Lallemand et al. (2015) found no clear consensus on quantitative versus qualitative approaches to UX, nor whether UX is an individual or social phenomenon, and importantly identified both a gap in the way UX is understood between industry and academia, and geographical differences in the way that it was related to marketing, usability, and emotion. This implies a need to understand communities of practice within their national context, as the differences may alter design outcomes.

A case study of the role of UX design professionals in Agile development by Bruun et al. (2018) used semi-structured interviews to gather data on industry practice in a single case company in Denmark. They found that UX professionals had a wide range of responsibilities, which made it more difficult to characterise their activities. The focus of their research was how well UX tasks integrated into Agile development, rather than the detail of those tasks, and they identified a need to better understand the relationship between customer centric and user centric needs as that may affect the willingness to fund UX activity. Our choice of open guiding questions for the workshop was intended to draw out information in a role-neutral way and give room for any such conflicting demands to emerge.

An interview study of practitioners attitudes to design methods, techniques, tools and processes by Gray (2016) identified that method use by practitioners departed from the projected use in the HCI literature, having an increased reliance on professional judgement and an appropriate mindset for tailoring the method to the specific problem. The lack of direct engagement between the research and practice communities was raised as a concern. Our choice of methods for this study was motivated by a desire to make participation in our study interesting and enjoyable to encourage further engagement.

The most challenging situation for our research question was expected to occur if non-Agile practices of design discovery were conducted in an environment tailored for Agile software development. A study addressing a matching problem by Kuusinen et al. (2016) used online surveys and Ketso workshops to ask practitioners about their experience of applying Agile methods in a non-Agile environment. The focus of their study was on the environment of practice rather than the nature of the practices themselves, but it showed the utility of the Ketso method with practitioners. Their analysis identified management buy-in and a supportive organisational culture and structures, the team environment itself, and effective application of practices as key themes.

No prior work was identified that specifically explores design discovery practice in the UK software industry. This study addresses that research gap.

3. THE RESEARCH STUDY

Our purpose in conducting this study was to gather information on current design discovery practice, and to better understand the context of that practice by exploring the objectives that practitioners aim to satisfy. Face-to-face workshops were chosen as the means of data collection to allow a free exchange of ideas between the participants and gain richer information than might be obtained by surveys. A Ketso (Tippett et al., 2007) workshop format was chosen to facilitate this. Ketso is a technique for engaging communities in discussion around specific topics. A Ketso session builds up a picture of participants' ideas written onto 'leaves' that are placed on 'branches' on a felt background (see Figure 1). This picture emerges through a structured discussion about the topic, in this case practices used during design discovery. This approach also fosters a safe environment where all participants are able to contribute equally without any one individual dominating the discussion. Ketso achieves this by combining individual idea generation and group discussion, structured by the workshop materials and by the guiding questions asked by the facilitator.

Two workshops were held. The first workshop ran on university premises in late June 2019 with three participants. The second session ran with a large retailer and was held in their offices in mid-January 2020, with nine participants in three groups of three. Both took approximately 90 minutes including set-up, briefing, and clear-up. Relevant participant characteristics are listed in Table 1. None of the participants had used Ketso before.

***Table 1**: Participant characteristics*

Id	Domain	Role or specialisms
1	*HCI research*	*Rapid prototyping*
2	*HCI research*	*Assistive technologies*
3	*Manufacturing*	*Design*
4	*Retail*	*User research*
5	*Retail*	*User research, Management*
6	*Retail*	*Design, Development*
7	*Retail*	*Design*
8	*Retail*	*User research*
9	*Retail*	*User research*
10	*Retail*	*Design*
11	*Retail*	*Design*
12	*Retail*	*Design, Management*

3.1 Methodology

3.1.1 Ontological and epistemological position

Our assumption is that the data gathered describes only how things seem to be and any conclusions drawn may only be applicable to the participating communities of practice, and we therefore take a bounded descriptive-relativist ontological position. The findings of thematic analysis are reflexively created as the research progresses, and our understanding of professional practices arising out of social interactions and sense-making is itself a social construct, so our epistemological position is pragmatic social constructionist with an interpretivist theoretical perspective (Moon and Blackman).

3.1.2 Research question and relevant themes

The research question and sub-questions can be stated as:

How do UX practitioners approach and perform discovery?

- What is done in practice?
- What would improve practice?
- What are the challenges?

Any pattern of response in the workshops that addressed these questions was taken as an emergent theme.

3.1.3 Scope, intent, and depth of meaning

The aim was to develop a rich description of the whole, rather than a detailed account of one aspect, and our intention was an inductive analysis linked closely to the data, not a theoretical model of

practice. The data corpus for this study implicitly included many hours of conversation with practitioners prior to the workshops, so although extracts for coding were identified at the explicit semantic level in the transcript, the themes under which they were grouped were necessarily based on our interpretative understanding of their latent meaning.

3.2 Procedure

3.2.1 Ketso general details

Participants were asked a guiding question, and asked to write their own ideas onto leaf shapes. The Ketso leaves are colour coded to represent the kinds of ideas that are wanted at that stage, and have a letter in the corner of the leaf for those without full colour vision. Only one kind of leaf was provided for each question. The standard Ketso conventions were used:

- goals or next steps – yellow (Y)
- what works well – brown (B)
- creative new ideas – green (G)
- challenges or barriers – grey (-)

Taking turns, they introduced and explained their ideas to the group, and the leaves were then placed on to a felt workspace. The felt has a space at the centre, from which narrow coloured strips radiate out, representing branches. Oval label shapes were used for a reminder of the overall question, placed in the centre of the felt, and for labelling the branches. Each leaf was placed either onto a new branch or onto an existing branch that they seem related to. After introducing their individual ideas, the group discussed them, and were able to add more ideas or move them around if they saw more relevant connections. The facilitator then asked the next question. At any stage, a collectively agreed label could be written and added to a branch. Part of the workspace from the first session is shown in Figure 1.

Figure 1: *Ketso felt workspace from the pilot session*

3.2.2 Recruitment and setup

Recruitment was by internal communication within the organisations involved once initial contact had been made. In the case of the large retail organisation, this was based around their internal community of practice and the workshop took place in one of their regular meeting slots, at which they were accustomed to trying out new methods.

Participants were given a brief general overview of the format of the workshop and the Ketso materials. More detailed information on the use of the materials was given as each guiding question was introduced by the facilitator. Around each Ketso workspace, three seated participants sat at each table. There was sufficient room for all to have spare leaves and writing space, without any having to view the text upside-down or from an uncomfortable angle. Five minutes were allocated for idea generation, and 10 minutes for group discussion, of each question. Each guiding question was supported with prompts for the kinds of things we would like them to consider, and written up on a poster in the list form shown.

3.2.3. Guiding questions

Question 1 (Yellow) asked about their success criteria. They were prompted to think about how they would recognise a ‘good' discovery session, what it should look like and feel like, and what the discovery should produce as an outcome or output.

Question 2 (Brown) asked what was currently working well. Based on the findings of Gray (2016), we explicitly prompted them to include mindset as one of the things to consider, in addition to methods and materials.

Question 3 (Green) asked what they would like to do differently if they could. Time was then allowed for reflection on which things mattered most, and how the ideas were related to each other.

Question 4 (Grey) asked what obstacles and challenges they had, and Question 5 (Green) asked how they might solve them.

Question 6 (Yellow) asked if they wanted to update their success criteria in the light of the discussion, adding any additional goals they felt were appropriate.

1. What does successful discovery

- look like?
- feel like?
- produce?

2. What works for you now

- mindsets?
- methods?
- materials?

3. What would you try with

- more time / people?
- more space?
- permission to fail?

4. What are the challenges

- behaviour?
- surprises?
- technology?

5. How might you solve them

- mindsets?
- methods?
- materials?

6. How is our vision of success

- any new goals?
- any new criteria?
- any new priorities?

3.2.2 Data collection and processing
The Ketso leaves have adhesive strips on the back that allow the whole felt workspace to be folded up and packed away for later transcription, with only minimal movement of the shapes against the background. To assist transcription, photographs of each workspace were taken with a smart phone before packing them away.

The content of each leaf was transcribed into a spreadsheet to capture the raw textual content before preparing a document for analysis, using the standard template from the ketso.com website. This was shared with the participants within 24 hours of the workshop, as a courtesy and for their own use should they wish to. The spreadsheet also captured which felt each idea was from, which branch it appeared on, and what type of leaf had been used.

3.3 Data analysis

3.3.1 Artefacts generated
To provide a permanent copy of each workspace, a digital version was created and checked against the photographs, with the exact text and the same relative positions of the leaves on each branch. A copy of this was provided back to the participants for their own use, accompanied by a reminder of the questions they had been asked, in a summary legend sheet. The Portable Document Format (PDF) copy of each workspace was used as the input document for coding and thematic analysis using the NVivo tool.

3.3.2 Thematic analysis
The ideas gathered at the workshops were analysed from the perspective of the framing questions, using the thematic analysis approach suggested by Braun and Clarke (2021). Only the text of the idea was used in the analysis, not information on which felt it came from, or the branch label that participants had applied to it. As such, all leaves were treated equally, and themes allowed to emerge from the text as a whole rather than from any structure imposed by the participants or implied by the guiding questions. The kind of leaf used was not generally taken as significant unless it helped distinguish a goal from a challenge.

Representative labels for the ideas were chosen by *in vivo* coding from the words used by the participants, or synthesised from the underlying concepts if their words were not sufficiently general, and then relabelled or merged as broader themes emerged from the data. Ranking of the themes, by the number of contributing participant groups and the number of textual references, was used to identify the most prominent ideas for the purposes of consistent presentation and communication. As the data gathered was insufficient to be considered representative of the wider community of practice no other significance should be attributed to the ordering.

3.4 Results

A total of 250 Ketso ‘leaves’ were completed by 12 workshop participants, of which 9 were UX practitioners from a large retail organisation, and 3 were academic staff with an equivalent background in product design or ethnographic research. These provided a total of 74 statements of current practice that participants considered to work well, 61 statements of aspirational practice, 64 statements of challenges to successful discovery, and 51 statements of what constitutes successful discovery.

We asked practitioners for their ideas on how a successful design discovery could be characterised. The analysis identified common terms of reference for challenges, constrained resources, attributes of success, and means of successful discovery. These were then further collected into key aspects of practice needed to meet operational goals in Table 3, organisational aspirations in Table 4, and organisational challenges in Table 5.

3.4.1 Coding examples
Phrases used by the participants were preferred as the initial representative of that idea. These were then progressively merged together under more general codes until the differences between their meanings had sufficient significance to keep them distinct. Examples of the approach to coding are given in Table 2. The codes ‘Enabling others' and ‘Empowering teams' were grouped with seven others into theme ‘Empowering'. Ideas about high quality artefacts to capture learning, research libraries, and sharing insights with other teams were coded as ‘Exchanging knowledge', which was grouped with code ‘User led' and four others into theme ‘Knowledge led'. The ‘Empowering' theme

was grouped with ‘Curiosity' and ‘Knowledge led' themes under the top level theme of ‘Organisational aspirations, as shown in Table 4.

Table 2: *Examples of coding*

Leaf text	Code
Training others how to do discovery	*Enabling others*
Help more people design and build for themselves	*Enabling others*
Empowered to say no	*Empowering teams*
Time and autonomy to get clear outcomes with team	*Empowering teams*
Produce lovely artefacts to show and save learnings	*Exchanging knowledge*
Research library	*Exchanging knowledge*
Share insights with other teams that may benefit	*Exchanging knowledge*
User led product direction	*User led*
Users being listened to	*User led*

3.4.2 Key aspects of practice

Key aspects of practice that emerged under the top level theme of operational goals are listed in Table 3, most prominent first. They were grouped under three themes: what they considered important in the methods that they used, what mindsets produced positive outcomes, and what constituted a positive outcome.

Methods were referred to in general terms, with no specific method being named. A preference for prototypes and experiments “allowing for randomness and unpredictability” was present in both sessions, but more pronounced in the academic setting. The retail organisation emphasised being certain “what to do next” and being able to assess whether to continue or stop. Being data driven by “using data to identify customer problems” and if necessary having “evidence to stop further progress” was as prominent as the use of ethnographic methods for “observation of users in real-world settings”. Goals of “having enough time” and making “efficient use of what you have available to you” were taken as a desire to be efficient, and the leaf type was used in that case to distinguish time as a goal from time as a challenge.

A practitioners mindset may affect the efficacy of the design discovery activities. Among the mindsets that were mentioned most were being confident and having “confidence in how to progress”, and “thinking laterally” to gain insights, and being inspired so that “there is a buzz around the success of the discovery”. An open attitude to “advertising challenges/progress” and “open sharing communication” was recognised as something that worked well, as was a purposeful mindset with an “emphasis on action/doing above all else'” and a “strong process”. An engaging mindset “involving others” and “sharing”, and a “collaborative mindset” where “the whole team has a shared understanding and has participated” were both identified as things that worked well.

Outcomes that were valued were that the problem should be understood, expressed for example as “the team understand the audience”, that this understanding be shared and aligned across the team so that the “team is on the same page regarding outcomes”, and that problem be defined and bounded so that they have a “clear scope for the next stage” and validated by “asking the right questions”. They wanted a detailed understanding that was “in-depth, not vague”, and some expressed a desire for data that could be visualised, for example by “displaying our work within our workspaces”.

Table 3: *Key aspects of practice*

Methods	*Experimenting, Certain, Data driven, Ethnographic, Efficient, Justified, Designer led, Human Centred, Iterative, Multi-disciplinary, Time-boxed*
Mindset	*Confident, Insightful, Inspired, Open, Purposeful, Engaging, Collaborative Honest, Alert, Curious, Empathetic, Flexible, Pragmatic*
Outcomes	*Understood, Validated, Aligned, Defined, Bounded, Visualised, Detailed*

3.4.3 Aspirational practices

Participants were asked what they would like to do if they were not constrained by the challenges they identified. The aspirational practice themes that emerged are given in Table 4. There was a strong theme of empowerment and autonomy, and both a desire to spend more time with stakeholders but also to “take stakeholder objectives out of the equation”. A desire to conduct both a broader and deeper discovery was expressed, to do more “in the wild” work and “have time to explore the whole ecosystem”.

As project teams move from periodic software delivery to something closer to a continuous delivery model, there will be a need for discovery activity to integrate with development in ways that better support that, such as the practices described by Torres (2021). One participant expressed an interest in “rolling discovery to explore new areas”, suggesting that user researchers might have other reasons for wanting continuous discovery.

The university session included participants with a physical, rather than software, product background. Their aspirations included “trying lots of new technology to consider solutions” and “loads of money and people for prototypes”.

No organisational challenge associated with the diversity of user research participants was raised, but an aspiration for “easy access to a diverse

audience" suggests that this might sometimes be a problem.

Table 4: *Aspirational practice*

Curiosity	*Deeper discovery, Creative freedom, Broader discovery, Continuous discovery, Solution feedback, Diverse participants*
Empowering	*Empowering teams, Collaborative working, Developing capability, Enabling others, User engagement, Organisational agility, Flexible schedule, Stakeholder engagement, Strengthening practice*
Knowledge	*Exchanging knowledge, Competitor analysis, Sharing best practice, Persistent knowledge, User led, Sharing understanding*

3.4.4 Organisational challenges

The challenges identified were wider ranging, with weaker themes, but prominent among them were internal and external communication issues, constraints on time, wrong mindset, and inefficient processes, as shown in Table 4. Time pressure was associated not just with deadlines, but also having "no time to collaborate". Recruitment of necessary expertise was noted as a problem for understanding complexity and a problem of timing as they could not "recruit fast enough". References to an unhelpful mindset, particularly "solutionising" and "solution-led thinking", were common. Equally prominent were references to inefficient processes related to governance and sign-off.

The solutions that the participants discussed for the challenges given in Table 5 are reflected predominantly in the aspirational practices listed in Table 5, but also to an extent in the key aspects listed in Table 3.

Table 5: *Organisational challenges*

Communication	*Lack of clarity, Internal communication, External communication*
Constrained resources (human)	*Time, Suitably qualified people, Sophisticated knowledge, Workload, Research participants, Limited scope*
Constrained resources (material)	*Funding, Equipment, Legacy equipment*
Behavioural obstacles	*Wrong mindset, Low engagement, Lack of foresight, Solution driven, Disruptive incentives, Fear, Hierarchy, Bias, Over-specification*
Process	*Inefficient processes, Low-value activity, Inconsistent approaches, Rigid processes*

4. DISCUSSION

4.1 Discovery practices

The research question asked in this study was

> How do UX practitioners approach and perform discovery?

This was refined in the questions that primed the thematic analysis, and the guiding questions that participants were asked, to three questions:

- What is done in practice?
- What would improve practice?
- What are the challenges?

In answer to these questions we found that methods were referred to in general terms, with no specific method being named. A preference for prototypes and experiments was more pronounced in the academic setting than the retail organisation, which emphasised knowing how to proceed. Practitioners valued understanding that was shared and aligned across the team, a problem that was defined and bounded in scope, and research outputs validated by input from multiple sources. There was preference for data that was detailed, not vague, and which could be visualised. A good discovery was recognised by feelings of inspiration and excitement in the team. The breadth of factors discussed indicates a diverse experience of discovery, and an ad-hoc definition of success with no widely shared criteria within the organisation

In the descriptions of the practices they aspired to, there was a strong theme of empowerment and autonomy, and interestingly both a desire to spend more time with stakeholders but also to be less constrained by their objectives. A desire to conduct both a broader and deeper discovery was expressed, which suggests that exercising greater autonomy and achieving the desired 'user-led' process might require a more time efficient approach.

One of the factors most frequently discussed was time pressure. A greater focus on customer value and agility in development (Clarke et al., 2018) implies similar demands for agility in user research and other discovery activities. The mention of inefficient processes in the organisational challenges discussed may also be related to a feeling of insufficient time. Currently, discovery and development are often separate streams of activity such as the dual-track approach described by Cagan (2018). Research is needed to identify the extent to which continuous discovery methods (Torres, 2021) have been adopted in practice and to assess whether they successfully - avoid- sharp peaks in demand.

An interesting omission from the data was vocabulary associated with rigour and challenge. This was missing from both of the sessions, and was not a point of difference between the participants with retail and academic backgrounds. If challenge is not considered an important part of discovery, that might be because it is more strongly associated with later stages of development.

4.2 Tailoring the workshop format

The standard Ketso pack assumes up to eight people per workspace, but our experience with the first session suggested this would be too many, so in order to limit the number of people around each table to three or four, for comfort and viewing angle, an additional workspace was purchased. The number of leaves written by the participating design professionals, who were experienced in similar activities if not with Ketso, was sufficient that freedom to arrange them as they wished might have been curtailed if we had not done so.

For a complete cycle of questions – starting with a definition of done, covering what works or does not, and revisiting our definition of success – a period of 90 minutes was barely sufficient to allow proper discussion. If the availability of meeting spaces is limited, the ease with which the felt workspace can be folded and packed up without disturbing the leaves could be exploited to hold a follow-up discussion session at another time or with the workspace mounted vertically on a convenient wall space rather than on a table.

The physical writing and manipulation of the leaves may help the thought process, so it would be interesting to compare the face-to-face use of the physical Ketso materials with a fully digital equivalent on a Miro board or any similar online workshop platform.

4.3 Use as a research output

Capturing the raw text of each leaf was relatively quick and easy. Producing a high quality digital version of the workspace for use as a research output was more time consuming. Companies are becoming more interested in building research libraries as part of the operationalisation of user research and design activity, referred to as ResearchOps and DesignOps respectively, as described in Metzler (2020) and Dörnenburg (2018) or Malouf et al. (2019). High quality research outputs was identified in post-workshop discussions as an important part of their practice by the retail organisation that hosted the second workshop, so the relative ease and quality of artefact generation may remain an important consideration regardless of other changes in working practices brought about by greater use of remote working.

4.4 Limitations

Limitations of the approach used were that we only ran two workshops with participants from two companies, so the sample size was small. However, as this was an explorative, inductive study that aimed to identify an initial set of thematic categories to characterise aspects of current practice, a small sample was appropriate. The written data collected on the Ketso leaves was brief, and we did not audio record the workshop conversations as we wanted to encourage a relaxed atmosphere. As a result of this a more detailed understanding of some aspects of the written comments was not captured.

Due to limited time, participants were not asked to explicitly rank the relative importance of the factors discussed. The prominence of themes reflects the content of the workshop discussions but taken alone may not capture an accurate picture of working practices.

We asked participants what a ‘good' discovery looks like, and captured their ideas about the criteria that should be applied to judge it, but we did not explicitly ask who should make that judgement.

5. CONCLUSIONS

This study investigated design discovery, the crucial information gathering and sense-making part of the design process, which was under-explored in the academic literature. Using the Ketso workshop format, 12 user researchers and designers from a university and a large retail organisation were invited to share ideas about their design discovery process and its challenges. Our thematic analysis of the workshop outputs showed that practitioners valued an empirical data-led approach, where they could have confidence in the coverage and validity of the data, and achieve a shared understanding of the user research findings across the organisation. Their aspirations for future practice focussed on greater depth and breadth of user research, team empowerment, and knowledge exchange. Key challenges to successful discovery included a lack of clarity in communication, constrained time and materials, and an inappropriate mindset. Research is needed to identify the extent to which continuous discovery methods (Torres, 2021) have been adopted in practice and to assess whether they successfully address these challenges.

The Ketso community engagement method was used in 90 minute workshops, and was found to be an efficient and effective means of asking questions about working methods and sharing ideas within a community of practice. For use with professional designers, groups of no more than three or four people are recommended, as an adaptation to participants skilled in idea generation.

6. ACKNOWLEDGEMENTS

Initial contact with the participating organisation was made at one of the regular UX community events in Manchester. The networking and research recruitment opportunities of these events have been greatly missed during the current pandemic.

REFERENCES

BRAUN, V. & CLARKE, V. 2021. Analysing Qualitative Data in Psychology. *In:* LYONS, E. & COYLE, A. (eds.) Third ed. London: Sage.

BROWN, D. 2017. *Practical Design Discovery,* New York, A Book Apart.

BRUUN, A., LARUSDOTTIR, M. K., NIELSEN, L., NIELSEN, P. A. & PERSSON, J. S. The Role of UX Professionals in Agile Development: A Case Study from Industry. Proceedings of the 10th Nordic Conference on Human-Computer Interaction, 2018. New York: ACM, 352-363.

CAGAN, M. 2018. *Inspired: How to Create Tech Products Customers Love,* New Jersey, Wiley.

CLARKE, P., O'CONNOR, R. V. & YILMAZ, M. In Search of the Origins and Enduring Impact of Agile Software Development. Proceedings of the 2018 International Conference on Software and System Process, 2018 New York. Association for Computing Machinery, 142–146.

DÖRNENBURG, E. 2018. The Path to DevOps. *IEEE Software,* 35**,** 71-75.

GOTHELF, J. & SEIDEN, J. 2016. *Lean UX,* Sebastopol, O'Reilley.

GOVERNMENT DIGITAL SERVICE. 2019. *How the Discovery Phase Works* [Online]. Service Manual. Available: https://www.gov.uk/service-manual/agile-delivery/how-the-discovery-phase-works [Accessed 10 May 2021].

GRAY, C. M. "It's More of a Mindset Than a Method": UX Practitioners' Conception of Design Methods. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2016. New York: ACM, 4044-4055.

GREGORY, P., BARROCA, L., TAYLOR, K., SALAH, D. & SHARP, H. Agile Challenges in Practice: A Thematic Analysis. *In:* LASSENIUS, C., TORGEIR, D. & PAASIVAARA, M., eds. Agile Processes in Software Engineering and Extreme Programming, 2015. Cham: Springer International Publishing, 64-80.

KNAPP, J., ZERATSKY, J. & KOWITZ, B. 2016. *Sprint,* London, Bantam Press.

KUUSINEN, K., GREGORY, P., SHARP, H. & BARROCA, L. Strategies for Doing Agile in a non-Agile Environment. Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, 2016. New York: ACM, 5:1-5:6.

LALLEMAND, C., GRONIER, G. & KOENIG, V. 2015. User experience: A concept without consensus? Exploring practitioners' perspectives through an international survey. *Computers in Human Behavior,* 43**,** 35 - 48.

MALOUF, D., BLACK, M., WHITEHEAD, C. & BATTLES, K. 2019. DesignOps Handbook. InVision.

MCGRATH, R. G. 2010. Business Models: A Discovery Driven Approach. *Long Range Planning,* 43**,** 247-261.

METZLER, B. 2020. Leveling Up Your Research and Research Operations: Strategies for Scale. *Ethnographic Praxis in Industry Conference Proceedings,* 2020**,** 203-217.

MOON, K. & BLACKMAN, D. 2014. A Guide to Understanding Social Science Research for Natural Scientists. *Conservation Biology,* 28**,** 1167-1177.

SALAH, D., PAIGE, R. & CAIRNS, P. A Practitioner Perspective on Integrating Agile and User Centred Design. Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014 - Sand, Sea and Sky - Holiday HCI, 2014. Swindon: BCS, 100–109.

SHEPHERD, D. A. & GRUBER, M. 2020. The Lean Startup Framework: Closing the Academic–Practitioner Divide. *Entrepreneurship Theory and Practice***,** 1-31.

SIM, G. R., ZAMAN, B. & HORTON, M. P. L. A Method Impact Assessment Framework for User Experience Evaluations with Children. Proceedings of British HCI Conference 2017, July 2017.

TIPPETT, J., HANDLEY, J. F. & RAVETZ, J. 2007. Meeting the challenges of sustainable development—A conceptual appraisal of a new methodology for participatory ecological planning. *Progress in Planning,* 67**,** 9 - 98.

TORRES, T. 2021. *Continuous Discovery Habits,* Bend, Oregon, Product Talk.

TRIEFLINGER, S., MÜNCH, J., HEISLER, B. & LANG, D. Essential Approaches to Dual-Track Agile: Results from a Grey Literature Review. *In:* KLOTINS, E. & WNUK, K., eds. Software Business, 2021. Cham: Springer International Publishing, 55-69.

Paper session 7: Education. Session chair: Parisa Saadati

Towards a digital syllable-based reading intervention: An interview study with second- graders

Philipp Schaper[1], Melissa Donnermann[1], Nicole Doser[1], Martina Lein[1], Anna Riedmann[1], Sophia C. Steinhaeusser[1], Panagiotis Karegeorgos[2], Bettina Müller[2], Tobias Richter[2], Birgit Lugrin[1]
Human-Computer-Interaction[1] and Educational Psychology[2]
Julius-Maximilians-University of Wuerzburg
Wuerzburg, Germany
{philipp.schaper, anna.riedmann, bettina.mueller, birgit.lugrin}@uni-wuerzburg.de

Reading is an essential ability and a cornerstone of education. However, learning to read can be challenging for children. To scaffold young learners, a number of reading interventions were developed, including a syllable-based approach in German, which has proven to be successful, but resource and time consuming through individual interaction by educators. To improve the reach of the reading intervention, we present the first step towards a digital intervention, following a human-centred design approach. In this contribution, we present the implementation of a digital prototype, developed with the feedback of expert evaluations, as well as an interview study with second-graders. The results of interviews with children showed that the app is suitable to be applied in the target age group, that children had fun using it and were motivated to further do so. The study also provides design implications for transferring an analogue concept into a digital application.

elementary education, technology supported learning, user study

1. INTRODUCTION

Around 6.2 million people in Germany cannot read properly (Grotlüschen & Riekmann, 2011). The origin of the inability or a deficiency in reading capabilities can often be traced back to the initial attempts to learn reading during primary school. This problem is reflected in around 15% of German fourth graders, who show deficits in extracting meaning from presented texts (Bos, Tarelli, Bremerich-Vos, & Schwippert, 2012). In the overall reading performance, German children showed a lower mean performance score relative to most other European countries (Hußmann et al., 2017). This is due a stagnating overall score for Germany, which has not improved significantly since 2001 in contrast to many other European countries. However, there is also high variability in the reading abilities, with an increase in children on the lowest level of competence to 6% from 3% in 2001.

Given the high transparency of the German language that allows reading words on a letter-by-letter basis, even poor readers can read words accurately. However, this process is resource demanding and prone to errors (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), and thus not suited for understanding the meaning of a text. To read words accurately and fast, readers need to learn to use units larger than single letters (i.e., syllables and morphemes) – a step most poor readers in German experience difficulties with. Due to the relevance of reading abilities for knowledge acquisition (Nagler, Lindberg, & Hasselhorn, 2017) as well as social and cultural activities (Naumann, Artelt, Schneider, & Stanat, 2010), problems in reading acquisition should be addressed as early and as widespread as possible.

A number of analogue interventions were designed for poor readers in primary school in different languages (e.g., Ecalle, Magnan, & Calmus, 2009; Heikkilä, Aro, Närhi, Westerholm, & Ahonen, 2013), including German (e.g., Müller, Richter, & Karageorgos, 2020). The latter is a syllable-based reading intervention to foster word reading skills for groups of second-graders. A digital version thereof could facilitate learning anytime anywhere, allow personalized help and potentially provide a fun approach of additional training. In this contribution, we first present a theoretical overview on reading acquisition and the reading intervention. Second, we developed a high-fidelity prototype that digitalizes a fraction of the syllable-based reading intervention, alongside with an interview study with second-graders that aims to explore whether a digital version could be employed with the target age group. Our contribution is twofold: (1) We demonstrate a systematic approach to transferring established concepts into a digital format by adapting a validated analogous training for mobile devices; (2) We derive relevant design implications based on a target group oriented evaluation of the application.

1.1. Theoretical background on reading acquisition

In general, the acquisition of reading skills on the word-, sentence- and text-level of young children varies within the development process. Children quickly improve their reading abilities during the early grades, whereas the degree of improvement is lower in third-graders and subsequent grades (Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Logan et al., 2013). Due to the trajectory of reading skills development during grade one through four (Foorman et al., 1998), early intervention is essential to train deficient reading abilities with the aim to return children to the path of normal reading development (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Müller, Otterbein-Gutsche, & Richter, 2018). Without intervention, children, who have poor reading skills in early grades, are highly likely to remain poor readers (Landerl & Wimmer, 2008).

As children progress in reading development , they should reach the *consolidated phase* as they are able to memorize a larger amount of sight words by extracting larger chunks of grapheme-phoneme connections, including syllables (Ehri, 1995, 2005). The more words can be read by orthographic representations like syllables the faster word recognition is possible. Because syllables are often used in multiple words, readers might profit from transfer effects. Syllable-based reading therefore qualifies as a training technique to gain knowledge which is transferable to unknown texts or words (Huemer, Aro, Landerl, & Lyytinen, 2010). However, poor readers in German often experience difficulties in using (sub)lexical units like syllables or morphemes to recognize words faster via orthographic decoding processes (Landerl & Wimmer, 2008). In consequence, their word recognition is slow and error prone which leads to difficulties in sentence- and text-based reading comprehension. Their trajectory of reading skill development is less steep compared to readers with good word recognition skills (Pfost, Hattie, Dörfler, & Artelt, 2014).

Using the capacity of syllables, Müller et al. (2018; 2020) developed a syllable-based reading intervention for second-graders. The intervention aims to enhance the word recognition processes by repeatedly reading and segmenting syllables using 24 different game-like training sessions. The word material within the manualized intervention was selected based on the 500 most frequent German syllables in texts typically read by 6-8 year old children (cf. database childLex, Schroeder, Würzner, Heister, Geyken, & Kliegl, 2015). The intervention aims to strengthen the mental representations of syllables and words consisting of these syllables to foster the accuracy and fluency of word recognition. Thus, the intervention should support poor readers to reach the consolidated phase instead of reading words letter-by-letter.

Müller et al. (2018; 2020) evaluated the reading intervention in the form of biweekly training sessions organized by specifically prepared trainers in addition to regular lessons. Participating children showed significant improvements regarding phonological and orthographic word recognition, hence being able to read words more quickly and accurately compared to same-skilled poor readers in the control condition. Participants also profited from enhanced reading comprehension due to transfer effects (Müller, Richter, & Karageorgos, 2020; for results in Grade 4 see Müller, Richter, Karageorgos, Krawietz, & Ennemoser, 2017). The results of the intervention highlight the value of early support in reading development to effectively improve reading skills (Klatte, Steinbrink, Bergström, & Lachmann, 2013; Tacke, 2005). However, conducting such an intervention in small groups is time and resource consuming. Additionally, even though group-based interventions have been shown to work, individual support for each learner would be beneficial (Groth, Hasko, Bruder, Kunze, & Schulte-Körne, 2013).

1.2. Related work on the development of app-based learning

Mobile apps have been introduced into the field of education for young learners for a wide range of topics ranging from science to language acquisition (e.g., Henderson & Yeow, 2012; Walter-Laager et al., 2017). Several studies have highlighted the positive effects of integrating tablet apps in educational contexts (e.g., Bastian & Aufenanger, 2017; Schoppek & Tulis, 2010), even compared to desktop applications (Sung, Chang, & Liu, 2016). There are also indications that mobile apps increase motivation (Su & Cheng, 2015; Tillmann & Bremer, 2017). Further advantages include access at any time, versatility, adaptability (Rossing, Miller, Cecil, & Stamper, 2012) and individual feedback to consolidate learning progress (Blok, Oostdam, Otter, & Overmaat, 2002).

Although a great amount of educational apps exists for young learners, quantity does not ensure quality of content (Papadakis, Kalogiannakis, & Zaranis, 2018). Haßler, Major, and Hennessy (2016) criticize that only a small subset of tablet apps is empirically tested, and that only a few studies adhere to minimum quality criteria, stressing the need for further empirical research. The general notion that adequately tested mobile apps might benefit learning and that introducing technology into the classroom can additionally enhance interest and motivation appears to be well supported (Hochberg, Kuhn, & Müller, 2018).

There are several apps with the overarching goal to practice and improve reading. For example, there are apps with picture books for beginners in reading

(e.g., Cahill & McGill-Franzen, 2013), or multimodal e-books to train children's reading skills (e.g., Morgan, 2013), indicating different possibilities to support reading acquisition. Focusing on spelling in German, there are similar issues with the apps in terms of quality (Fleischhauer, Schledjewski, & Grosche, 2017). To date, no evidence-based app using a syllable-based approach with focus on orthographic and phonological processes is available.

The contrast between the large amount of reading apps available and the lack of using established and evaluated approaches highlights a need for a scientifically driven development process (Hirsh-Pasek et al., 2015). In order to obtain the benefits of mobile learning, the app needs to be designed focusing on its usability and handling by children. Comfort and ease of use are also important aspects. When adapting an established reading intervention from an analogous course-based format to a mobile app, several aspects have to be considered. For example, motivational aspects and self-directed learning become more relevant, because users interact with the app by themselves, and potentially have no external guidance (Ericsson, Krampe, & Tesch-Römer, 1993). A mobile app should promote self-directed learning on different levels of difficulties in pupils, similar to the benefits reported for technology-rich learning environments (Fahnoe & Mishra, 2013; Rashid & Asghar, 2016; Rossing et al., 2012; Underwood, Luckin, & Winters, 2012).

1.3. Contribution

Acquiring recognition and understanding of written words is one of the main educational goals in primary school. For poor readers learning to read in German, a syllable-based intervention has been shown to positively influence word reading skills and, in consequence, text-based reading comprehension (Müller, Richter, & Karageorgos, 2020; Müller et al., 2017). However, the analogue application of this approach is time and resource consuming, as many of the conventional interventions highly rely on individual interaction by educators to convey the learning material. A digital version of this intervention for reading acquisition could potentially overcome these limitations. However, the transfer of such an evidence-based face-to-face intervention into a digital version should be conducted with a human-centred design approach. As noted by Görgen, Huemer, Schulte-Körne, and Moll (2020), until recently the focus for development for children with reading problems has been on analogue trainings with experts. To the best of our knowledge, even though steps towards digital reading trainings have been made, they are not set up as transferred version of a tested analogue training. Even though steps in this direction have been taken by comparing pen-and-paper and digital content in math education (Maertens, Vandewaetere, Cornillie, & Desmet, 2014).

As a first step, this study addresses to which degree such analogue teaching material, which is originally presented by an instructor, can be transferred into a digital version. Perceived usability and enjoyment regarding the learning material might be crucial in determining the frequency of use of the digital version, both by children at home and teachers in their classes. Only if these potential motivational obstacles are overcome and the underlying principles of the intervention can be conveyed, a subsequent larger investigation on the actual learning outcome in the target group can be considered.

We therefore present the first step of a human-centred design approach to implement a digital version of the reading intervention by Müller, Richter, and Otterbein-Gutsche (2020), providing an operational high-fidelity prototype that integrates expert feedback from educational psychologists and primary school teachers as well as feedback from second-graders, the target group of the reading intervention. Our work demonstrates a systematic approach to transferring an analogue concept into a digital application and design implications derived from this process.

It is important to keep in mind that the current state of the app is not yet a complete digital version of the original intervention, for which similarly positive results on learning outcomes may be expected. However, the prototype is intended as an elaborated proof of concept on whether the intervention has the potential to be developed into a mobile, easy-to-use and motivating application, which is accepted by pupils and teachers alike and able to improve reading abilities, based on empirically tested scientific material. To this end, we addressed the following research questions:

RQ1: Can the central elements of the analogue syllable-based reading intervention be successfully transferred into a digital version?

RQ2: Are second-graders motivated to use the app again?

RQ3: Do the second-graders enjoy using the app?

2. PROTOTYPE OF THE READING APP

For the prototype, three educational games used in the reading intervention (Müller, Richter, & Otterbein-Gutsche, 2020) were chosen for implementation in Android Studio (Google, 2019). First, a selection of games were implemented as low-fidelity prototypes using Axure (Axure Software Solutions Inc., 2019). Second, three of these prototypes were chosen for implementation as an Android app, based on their

ability to capture essential aspects of the reading intervention such as the ability to separate words into syllables or the acquisition of syllable rhythm and syllable classification.

2.1. Analysis of requirements

By digitalizing the complete syllable-based intervention, it could potentially benefit from the advantages of technology-based tools (Biancarosa & Griffiths, 2012; Sung et al., 2016). A digital version could also avoid the logistic drawbacks of the analogue intervention, which requires specifically prepared teachers. Furthermore, as the app is used by each child individually, the pace for the training is not affected by the group (Groth et al., 2013) and an adaptive training at the child's individual reading level is possible.

We identify four main criteria for our app based on relevant requirements for mobile education applications for children as proposed by Mkpojiogu, Hussain, and Hassan (2018) and the PACMAD (People At the Centre of Mobile Application Development) usability model by Harrison, Flood, and Duce (2013): (1) efficiency in terms of adequate and accurate feedback, (2) effectiveness with regard to the comprehensibility of content and navigation elements, (3) the self-descriptiveness of the individual games and the app overall (learnability) and (4) child-friendly design to assure user satisfaction and reduce cognitive load. Additionally, we focused on the correctness of content. These requirements have to be considered for the adaption of the analogue version to a digital version, which is closely based on the original training while being extended by meaningful modifications, such as an immediate feedback mechanism, to fully exploit the potential of a digital adaption. For subsequent development, further requirements such as adapting the app's overall difficulty level to the children's individual skills requires further analysis of user data, e.g. by collecting performance data.

The correctness of the content is addressed by a direct transfer from the analogue reading intervention to the digital version, and a feedback loop via an expert evaluation. The four requirements for educational apps for children (Mkpojiogu et al., 2018) are dependent on design choices during the development process. The crucial idea behind these requirements is that second-graders, especially with reading deficiency, should be able to successfully interact with the app. Therefore, vivid explanatory videos that address both the navigation within the app, as well as the content and procedures within each individual game are integrated. Additionally, visual and acoustic feedback are used throughout the interaction using simple, consistent sounds with the main purpose to indicate if a given solution to a task was right or wrong (Mejtoft, Lindberg, Söderström, & Mårell-Olsson, 2017). By implementing colorful and expressive graphics, the app is designed for the target age group of the reading intervention.

Finally, because scientifically developed technologies are often not integrated into everyday classrooms (Scherer, Tondeur, Siddiq, & Baran, 2018), the to be developed app had to be accepted not only by experts, but also by pupils as well as their teachers to be of practical use.

2.2. Main concept of the digital intervention

The main menu of the app is depicted as a map of an island structured in different areas represented by distinctive symbols corresponding to the three respective games (flowers, lighthouse, and school). After selecting one of the games, the user can choose between three difficulty levels. Completing a game in one of the difficulties returns the user to the difficulty selection screen.

Within each difficulty level, five tasks are presented sequentially, with a progress bar illustrating the amount of finished tasks per level. Each task is set to be worked on in an answer-until-correct style (Blair, 2013; Epstein, Epstein, & Brosvic, 2001), allowing the learner to reach the correct solution eventually. Users are also allowed to abandon the currently played level and return to the difficulty selection screen.

The design of the buttons was kept simple and comprehensible. The difficulty selection screen for each respective game as well as the levels themselves use an image of the previous app layer in the upper left corner as a button to return to this layer. To ensure proper understanding regarding the functionality of the app, tutorial videos are displayed when starting the app for the first time and before each game. Each tutorial video can be repeated in each game by pressing a button labeled with a question mark.

2.3. Implementation of the selected games

The first implemented task is 'Reading with Arcs' in the analogue reading intervention. Children are instructed to mark the syllables of words with arcs and thereby read the words syllable-by-syllable. In the analogous intervention, syllables are marked with arcs while reading a word aloud with the whole class, as well as silently during individual work. The aim is to make syllables the salient unit of processing instead of reading words letter by letter (Müller, Richter, & Karageorgos, 2020).

The game 'Reading with Arcs' from the analogue intervention was adapted and called 'Flower Meadow' and is associated with the flower button in the selection screen. Similar to the analogue version of the game, users have to hyphenate words into syllables. Separation of the words is represented by the flight path of a bee. Every letter of the respective

word is placed on a separate flower (Figure 1). Users separate the syllables by dragging the icon of a bee to the final letter of the current syllable within the word. The number of syllables per word varies according to the level difficulty of the game. Visual and audio feedback is used to indicate a right or wrong answer. Once users correctly hyphenate a word, their progress is acknowledged and they proceed with the next word in the level. Difficulty is adapted by inclusion of longer words in higher levels. Once the progress bar in the level is filled, the user is informed about passing the game in the respective difficulty and allowed to return to the difficulty selection screen.

Figure 1: *The game 'Flower Meadow' during gameplay. (The letters show the word Dusche / shower).*

The second selected game is 'Vowel and Consonant', which serves primarily to teach the composition of a syllable and the distinction between vowels and consonants. The original version of the learning exercise describes each syllable as a boat, which has exactly one captain but several sailors. The vowel, or diphthong, is the captain of the boat, the consonants are the sailors that the children must recognize. In German each syllable contains a vowel or diphthong as its nucleus, therefore identifying the vowel or diphthong can help children to distinguish the syllables within a word. This general rule is introduced in the third session of the analogous intervention and is trained in every session afterwards. Recognizing individual syllables and their structure is a fundamental requirement for the differentiation of syllables and thus the reading of whole words.

In the digital version 'Vowel and Consonant' was adapted and renamed as 'Sailor Game' and is associated with the lighthouse button in the selection screen. The digital version visualizes the explanation used in the reading intervention of each syllable as a boat with a captain by dividing a word and placing its syllables in separate boats, initially representing each letter as a sailor (Figure 2). Users have to recognize the vowels, vowel mutations or diphthongs as the captain of each syllable of the depicted word. The goal for each level is to find all captains by touching the vowels and diphthongs. Positive feedback is shown by highlighting the correctly selected figures with a green border and changing the sailor into a captain, negative feedback by a red shading, and a shake animation. Furthermore, the facial expression of the selected sailor changes accordingly as well as the auditory feedback. The difficulty of each level varies due to the inclusion of monosyllabic words in difficulty level one, progressively to three-syllable words in difficulty level three. Once the progress bar in the level is filled, the user is informed about passing the game in the respective difficulty and allowed to return to the difficulty selection screen.

Figure 2: *The game 'Sailor Game' during gameplay. (The letters show the word besser / better).*

The third game 'Syllable Salad' occurs first in the sixth session of the reading intervention. In the original version of the intervention, the child receives an envelope, which contains cards with several separate syllables. The task is to bring these syllables in the correct order to merge them into words. In the digital version, 'Syllable Salad' is represented as a school building on the overview map. Users rearrange individual syllables to form a complete word. To this end, the syllables of a word are displayed randomly scattered on a blackboard (Figure 3). The task is to place the syllables in the right order by using a drag and drop mechanic. Positive and negative feedback is communicated via sound output. The individual levels vary in their degree of difficulty in terms of the number of syllables. In difficulty level one, two-syllable words are presented, while difficulty level two consists of three-syllable and difficulty level three of four-syllable words. Once the progress bar in the level is filled, the user is informed about passing the game in the respective difficulty and allowed to return to the difficulty selection screen.

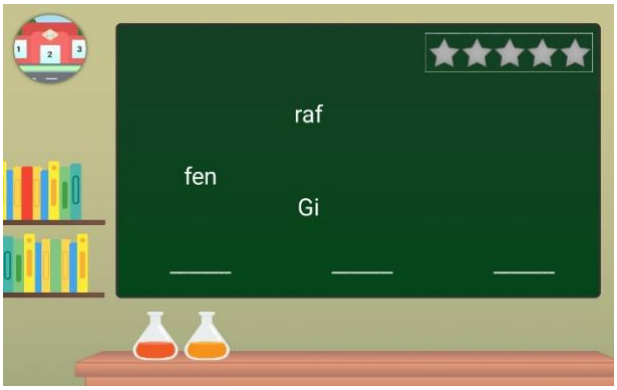


Figure 3: *The game 'Syllable Salad' during gameplay. (The letters show the word Giraffen / giraffes).*

2.4. Expert review

In order to examine whether the four requirements for the app (efficiency, effectiveness, learnability and child-friendly design to assure user satisfaction and reduce cognitive load) have been met by our implementation, and prior to conducting a study with second-graders, the prototype was evaluated by experts. Three researchers (one professor, 2 post-docs) at the chair of educational psychology separately used and commented on the app. In individual interviews, the experts first watched and commented the instruction videos. Afterwards, they were invited to use the app without time limit and to freely navigate within the app. During interaction with the app, the experts gave oral feedback, which was recorded. After they finished operating with the app, additional half-structured interviews were conducted. The questions focused on possible difficulties that could arise during the children's app usage, issues the pupils could need assistance with, and the adequacy of the implemented feedback. They were also asked to provide feedback for the procedure of the study based on their expertise, and to comment on the interview guideline for half-structured interviews, planned to be conducted with the children in the subsequent study.

Regarding the instruction videos, the experts mainly reported potential problems with the wording of the conveyed information and the feedback. The experts provided recommendations for the graphical user interface, especially for the main menu's design, the font size, as well as the font type within the games. They further suggested to include an instruction video to explain the general handling of the app. Finally, the usability of the game 'Syllable Salad' was criticized. All expert feedback was taken into account and the app was adapted accordingly.

After implementation of the expert feedback, we presented the app to the class teacher of the pupils participating in the study, as well as an additional primary school teacher. Both explicitly confirmed the app's and interview's suitability for the evaluation.

3. USER TESTING

We conducted a study with a school class of second-graders who interacted with the app in a controlled setting and took part in a subsequent interview. The main aims were to test if the central aspects of the analogue syllable-based reading intervention can be successfully transferred into a digital version (RQ1), whether second-graders rate the app as motivating (RQ2) and whether second-graders enjoy using the app (RQ3). For future adaptions, we were interested in the performance of the children using the app and therefore measured errors and time spent on specific tasks.

3.1. Participants

The study was conducted in a primary school in Germany. In total, 17 children with written consent by their legal guardians used the app. The age ranged from seven to nine years (M = 7.65, SD = 0.61). Twelve girls with a mean age of 7.75 years (SD = 0.62) and five boys with a mean age of 7.40 years (SD = 0.55) participated in the survey. According to the class teacher, the reading abilities of the participating children were heterogeneous.

3.2. Design and procedure

The study was separated into two parts for each participant, with two investigators assigned to one child. The sessions were held individually and were set to not exceed 40 minutes.

First, each pupil was invited to interact with the app presented on a tablet, playing each game for approximately five minutes, using the same sequence of the three games, 'Flower Meadow', 'Sailor Game' and ending with 'Syllable Salad'. Afterwards, each child could freely use the app for up to five more minutes. The screen capture software 'DU Recorder' (DU-Apps-Studio, 2018) was used to record the complete interaction with the app. Questions asked during the interaction and the children's behaviour were recorded manually in writing, including instances in which the child needed assistance.

Second, a half-structured qualitative interview was conducted based on the interview guidelines discussed with the experts. One of the investigators asked questions and interacted with the participant while the other researcher noted answers and behaviour. The interview consisted of 17 open-text questions with child-friendly wording, which concerned demography, including age and gender, experience with tablet usage, joy of use, perceived learning outcome, basic understanding of the app's handling, the app's design and the design of the individual games. The pupils were asked which of the games they perceived to be hardest and easiest respectively, which game they liked best and which game they wanted to play again in the future. There

were also open questions for positive and negative feedback on each individual game. Finally, the pupils were asked for general feedback and were invited to give additional suggestions.

4. RESULTS

The interviews were evaluated using the data and text analysis software MAXQDA 2018 (VERBI Software, 2018). Statistical tests were conducted using SPSS Version 25.

4.1. Screen-capture data

Due to technical problems, only 16 screen-capture videos could be analysed. Each video was evaluated independently by two investigators. The following criteria were extracted from each screen capture: (1) number of accomplished levels within five minutes, (2) time per level, (3) number of content-related errors, (4) number of interface-related errors and (5) choice of games during free interaction. Errors were separated into two categories to differentiate between problems with the interaction (interface-related), such as trying to interact with a non-interactive object, and problems with the reading task (content-related), such as selecting a wrong answer (Table 1).

Table 1: *Time per level, content-related and interface-related errors for all three games and their respective levels.*

		Time per level		**Number of content-related errors**		**Number of interface-related errors**	
		Mean (SD)	*Min - Max*	*Mean (SD)*	*Min - Max*	*Mean (SD)*	*Min - Max*
Flower Meadow	*Level 1 (n=16)*	1.32 (0.38)	0.53-2.59	2.88 (0.38)	0-5	2.69 (0.07)	0-11
	Level 2 (n=15)	1.29 (0.40)	0.50-3.39	4.87 (0.60)	0-5	1.60 (0.20)	0-3
	Level 3 (n=8)	1.26 (0.18)	0.59-1.58	5.11 (0.45)	0-4	0.56 (0.19)	0-2
Sailor Game	*Level 1 (n=16)*	0.57 (0.40)	0.16-2.40	2.69 (0.24)	0-3	0	0
	Level 2 (n=16)	0.40 (0.16)	0.18-1.19	1.56 (0.09)	0-3	0	0
	Level 3 (n=15)	0.46 (0.12)	0.25-1.18	1.40 (0.08)	0-3	0	0
Syllable Salad	*Level 1 (n=16)*	0.44 (0.07)	0.35-1.03	0.69 (0.13)	0-2	1.06 (0.15)	0-4
	Level 2 (n=16)	1.23 (0.28)	0.42-2.29	3.81 (0.55)	0-7	0.69 (0.07)	0-2
	Level 3 (n=15)	2.04 (0.25)	1.25-2.54	8.53 (0.93)	0-8	1.43 (0.19)	0-4

Note: Standard deviations are in parentheses.

Interface-related errors should be as low as possible while content-related errors are based on the ability of the learner and are not necessarily zero when confronted with more difficult tasks. Each child completed all levels of ‘Syllable Salad’ as well as all levels of ‘Sailor Game’, except for one participant who did not finish the third level before the five minutes had elapsed. In ‘Flower Meadow’, each participant finished the first level, but only nine children completed the third level within the five minutes.

The mean time spent on each level of ‘Flower Meadow’ was 01:29 minutes, for ‘Sailor Game’ it was 00:48 minutes per level and for ‘Syllable Salad’ the mean duration was 01:23 minutes per level. Only for the game ‘Syllable Salad’, the mean time increased with difficulty of the levels. In ‘Flower Meadow’ the average time per level was nearly the same for all three levels while in the sailor game the first level took longest (Figure 4).

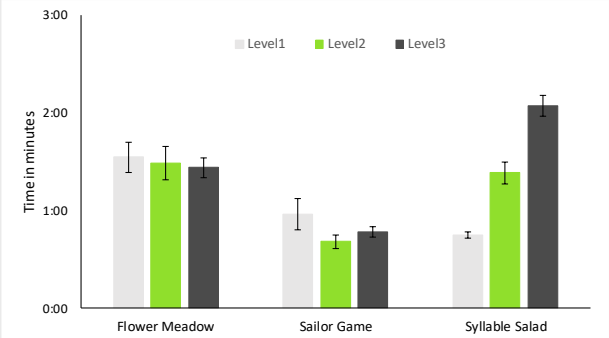


Figure 4: *Average time required for each completed level within each game. Error bars represent SEs.*

To assess content und interface-related errors we calculated the mean number of errors, only for completed levels, within each game, resulting in a decrease in participants in higher levels (see Figure 5).

Descriptively, the ‘Sailor Game‘ had the lowest number of content-related errors overall, followed by ‘Flower Meadow‘ and finally ‘Syllable Salad’ with the highest number of content errors. It is noteworthy that ‘Sailor Game‘ had the highest error rate in the first level whereas in ‘Flower Meadow’ and ‘Syllable Salad‘ the content-related errors increased with the difficulty of the levels. Concerning the interface-related errors, ‘Sailor Game’ had no interface-related errors at all. ‘Syllable Salad’ had fewer interface-related errors compared to ‘Flower Meadow’.

When being allowed to play with the app freely, two of the children were not interested in more interaction with the app, and for three participants data was not recorded during their interaction. The remaining pupils (n = 13) played one or multiple games during their free interaction. ‘Flower Meadow’ was played by ten children, ‘Sailor Game’ by eight children, and ‘Syllable Salad’ was chosen by six children.

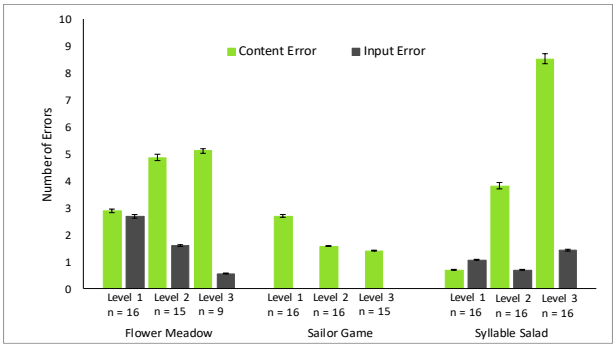


Figure 5: *Mean content errors and mean interface-related errors per child and level. Error bars represent SEs.*

4.2. Interview data

All interview data was separated into segments, which were assigned a category by a group of three raters based on majority. The same three raters assigned categories for all interviews.

Regarding the app in general, all children reported having fun while using the app and separately claimed that they liked the appearance of the app. Concerning the overall interaction, 13 participants stated that the handling of the app was clear and comprehensible to them.

When asked about the enjoyment of each game separately, all children affirmed that they enjoyed playing the ‘Sailor Game’. The pupils gave different reasons why they enjoyed the game, such as the game being easy to solve, easy to use, funny, that they liked the sailor theme and the visual feedback. For ‘Flower Meadow’, 15 children stated that they enjoyed the game, with the main reason of having fun playing the game. For ‘Syllable Salad’ also 15 children affirmed that they enjoyed playing the game and stated having fun and handling the game easily. Additionally, four children also mentioned the learning effect and the challenge the game posed.

Regarding the perceived difficulty, six children chose ‘Syllable Salad’ as the most difficult game, while ‘Flower Meadow’ and ‘Sailor Game’ were mentioned four times each. As the easiest game, ‘Sailor Game’ was chosen seven times, ‘Flower Meadow’ six times and ‘Syllable Salad’ three times.

When the participants had to decide on their favourite game within the app, six of the children chose ‘Flower Meadow’, followed by five children selecting ‘Syllable Salad’ and four of the children choosing ‘Sailor Game’. Two children had no favourite game. The children stated that their choice was based on the easy handling, easy solvability, fun and overall appearance.

When asked whether they would like to use the app again, all children agreed. However, ten children were only interested to replay certain games: 'Syllable Salad' as well as 'Sailor Game' were mentioned seven times, 'Flower Meadow' was named three times.

Interestingly, 14 children stated that they have learned something while playing with the app. Improving in splitting syllables was mentioned by eleven children, four stated that they improved their understanding of words and vowels, and two said to have improved in reading.

When asked for general feedback on the app, four of the children emphasized the graphics, three mentioned having fun and two commented on the perceived learning. Additionally, three pupils mentioned the simple handling and two enjoyed the challenge of the tasks. Regarding negative aspects, one of the children did not like the type of visual feedback and one noted that the difficulty might not be suited based on individual abilities.

Last, the pupils were asked for suggestions on how to improve the app. In general, improvement of graphics was mentioned four times. Additionally three of the children asked to implement more games and levels. Most suggestions for improvement occurred on 'Syllable Salad'. Individual suggestions included to use funny words, increase difficulty and use different graphics. In total, having fun while using the app was mentioned 31 times across all interviews. The easy interaction was named 15 times and the perceived knowledge gain ten times.

5. DISCUSSION

Concerning RQ1, the evaluation by psychological experts, teachers and the study demonstrated that the selected games of the analogue training can be digitalized. The pupils were overall able to independently work with the app based on the integrated instructions. However, the results have to be considered in detail. The quantitative data demonstrated substantial variability within the participating children in the abilities needed for the content of the app. There was also high variability in the interface-related errors for 'Flower Meadow' and 'Syllable Salad'. However, content-related errors were more numerous than interface-related errors in all three games. Only in the 'Sailor Game' no interface-related errors occurred, indicating the need to improve the interaction within both other games. For RQ2 and RQ3 we suggest that the results, especially the interviews, might have been impacted by the novelty of the games itself, especially relative to day-to-day school activities.

The interaction with the app itself should be unaffected in terms of content-related and interface-related errors and we still consider the information gained in the interviews as valuable insights into the target group. The overall reception of the app and the three implemented games was very positive. Given the fact that the games are clearly focused on learning, the interest of all participating children to use the app again is an encouraging sign, supporting the notion that the app is perceived as motivating, even though it should be treated with caution. Furthermore, the heterogeneous choice as favourite game can be considered as an indicator that the games are comparably enjoyable, which is supported by how often the children reported having fun, which supports RQ3.

5.1. Lessons learned

However, the performance of the children within the app in terms of time per level, content-related and interface-related errors indicates room for improvements. The time per level, as well as content-related errors increased in higher levels in 'Syllable Salad', with only few interface-related errors in each level. This suggests that the planned increase in difficulty across levels worked and shows that there appear to be no significant problems in using the game.

For 'Sailor Game' the mean time per level is similar across levels and the low mean number of content-related errors is even decreasing in higher levels. There were no interface-related errors at all for this game. 'Sailor Game' can therefore be considered as a relatively easy game in terms of content, which is also easy to use, as supported by the interview data.

Concerning 'Flower Meadow', the mean time per level is also constant across levels and content-related errors increased slightly in higher levels. Both measures can be considered as an indicator that the planned increase in difficulty across levels was noticeable. However, the number of interface-related errors in 'Flower Meadow' is descriptively higher compared to both other games, especially in the first level. We suggest that, even though the game appears to be not very difficult in terms of content, the current mode of interaction is not ideal, which results in increased interface-related errors. The interaction should be improved, especially because the games were selected to be suited as introduction for the app and therefore should be easy to interact with. Furthermore, these interface-related errors might be perceived as content-related errors by the children, potentially resulting in confusion about the correct solution.

The difficulty of 'Syllable Salad' and 'Flower Meadow' measured in errors matches the subjectively perceived difficulty of the games as indicated by the choices for easiest and most difficult game. Interestingly the larger amount of input-related errors in 'Flower Meadow' did not appear to have influenced this assessment. Furthermore,

considering the choice of favourite game, ‘Flower Meadow’ was chosen most often. The input-errors might therefore not be regarded as a critical problem for the children, this is most likely the case because errors were not costly and the correct input could be executed without delay.

When the children could freely interact with the app, half of the children chose the levels of the games they did not manage to finish before. This behaviour can be interpreted as an indication that the children were motivated to finish all levels. The implemented aspects of gamified elements in the learning environment such as the progress bars within each level and the increasing difficulty across levels might have incentivised the children to rise up to these challenges. In summary, the degree of difficulty appeared appropriate and motivating for the children and the overall interaction with the app worked well.

Considering the initial research aim of adapting an analogue reading intervention to a mobile app, several aspects can be noted. Foremost, based on the initial prototype and the reactions within the study, it can be assumed that central elements of the intervention can be implemented into an app (RQ1). Furthermore, the results of the interviews illustrate the children's motivation to use the app again, (RQ2). The interviews also show an overall enjoyment of the children using the app (RQ3). The results should be considered in light of a relatively small group of participants in a situation that is not equivalent to the planned context of use. We still regard these preliminary indications for future development as highly promising.

The combined approach of capturing and analysing the interactions with the games and the use of the interviews allowed to identify problems as well as potentials. This allows us to address the problem of input-errors in a more differentiated way, seeing that even though minimizing these errors is an obvious task in development, the consideration of a design, which minimizes their impact is also worthwhile for subsequent versions.

5.2. Design implications

Based on the results of the interviews and analysis of the screen-capture data we derived several concrete improvements for the app that might also serve as general design implications for app development in this area. Regarding the app’s user interface it is noticeable, that the use of text requires a font that is easy to read and therefore more suitable for beginning readers. Furthermore, the function of the buttons was not always clear. Although the buttons were already designed to be noticeable in size and contrast, it can be generally noted that app developers should focus more on the layout of critical user interface elements, such as buttons. This includes their self-explanatory presentation in the app and intuitive use. This is in line with the need for help by some children throughout the study. Therefore, the tutorials and help functions have to be extended to improve understanding. These observations indicate that especially children might profit from a more interactive way of explaining the core app functions, including active participation and exercises.

Furthermore, our results highlight the need of a stepwise and differentiated approach for transferring an analogue concept into a digital application. The selected requirements (1) efficiency in terms of adequate and accurate feedback, (2) effectiveness with regard to the comprehensibility of content and navigation elements, (3) the self-descriptiveness of the individual games and the app overall (learnability) and (4) child-friendly design to assure user satisfaction and reduce cognitive load can only be addressed by an iterative process including experts and the target group of children. Expert feedback alone, which is often used in form of heuristic evaluations has clearly not been sufficient to finalize the app as demonstrated by the study. However, the study with the second-graders by itself might have been similarly problematic because central problems such as the approach to the instructions and potential difficulties with the initial font had to be identified to actually allow for meaningful interaction with the app.

6. CONCLUSION

For this contribution, we used a human-centred design approach to develop a prototype of a mobile learning app based on an analogue syllable-based reading intervention (Müller, Richter, & Otterbein-Gutsche, 2020). Such a digital version could help to address the need for a more accessible reading intervention for children with deficient reading abilities.

The preliminary prototype was subject to an evaluation by experts and was subsequently revised. A study was conducted with a class of second-graders, who played with the app and were interviewed about the interaction. In general, the participating children enjoyed using the app and were motivated for continued use. The study demonstrated that core aspects of the analogue reading intervention can be adapted into a digital version. Results of the study generated valuable insights to improve the current state of the prototype. The feedback from the target age group can help guide future development of the app, e.g. on how to improve the introduction tutorials, therefore ensuring the quality of the developed app (Papadakis et al., 2018). These insights also provide implications regarding the development and design of educational apps for children.

Engaging a human-centred design approach allowed to incorporate the perspective of theoretical

experts (i.e., pedagogical psychologists), practical experts (i.e., primary school teachers) and users (i.e., second-graders). This enabled us to iteratively improve the prototype, establish a solid foundation for subsequent development of the prototype into a complete app, and to present this development approach as insight for similar endeavours.

The participating children also asked for extension of the levels and the app itself, which is less a needed improvement and more an indication that there is demand for the adaption of the reading intervention by Müller et al. (2018).

In the long term, the learning outcome of the app needs to be examined with pre and post tests for children with different reading abilities in a large sample. Because the underlying intervention has been shown to be effective (Müller, Richter, & Karageorgos, 2020; Müller et al., 2017) in a long term study, the effect of such a short interaction as in the study is only the first step in a comprehensive evaluation of the digital adaption. The goals during development were focused on our research questions: the proof of concept, motivation in the target group and interest in use of the app. Our goal is therefore to scientifically establish a usable and well thought out app, which adequately represents and meaningfully expands the analogue intervention, and also test this improved version in terms of learning success on a large scale (Haßler et al., 2016).

7. ACKNOWLEDGEMENTS

This work was funded by the German Federal Ministry of Education and Research (BMBF) under the grand agreement MobiLe (03VP07080).

8. REFERENCES

Axure Software Solutions Inc. (2019). Axure. San Diego, USA: Axure Software Solutions Inc.

Bastian, J., & Aufenanger, S. (2017). *Tablets in Schule und Unterricht [Tablet use in school and teaching]*. Wiesbaden: Springer.

Biancarosa, G., & Griffiths, G. G. (2012). Technology tools to support reading in the digital age. *The Future of Children, 22*(2), 139-160.

Blair, K. P. (2013). *Learning in critter corral: evaluating three kinds of feedback in a preschool math app.* Paper presented at the Proceedings of the 12th International Conference on Interaction Design and Children.

Blok, H., Oostdam, R., Otter, M. E., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of educational research, 72*(1), 101-130.

Bos, W., Tarelli, I., Bremerich-Vos, A., & Schwippert, K. (2012). *IGLU 2011 Lesekompetenzen von Grundschulkindern im internationalen Vergleich [IGLU 2011: International Comparison of Reading Skills in Primary School]*. Münster: Waxmann.

Cahill, M., & McGill-Franzen, A. (2013). Selecting “app” ealing and “app” ropriate book apps for beginning readers. *The reading teacher, 67*(1), 30-39.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review, 108*(1), 204.

DU-Apps-Studio. (2018). DU Recorder 1.7.5.5.

Ecalle, J., Magnan, A., & Calmus, C. (2009). Lasting effects on literacy skills with a computer-assisted learning using syllabic units in low-progress readers. *Computers & Education, 52*(3), 554-561.

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading, 18*(2), 116-125.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of reading, 9*(2), 167-188.

Epstein, M. L., Epstein, B. B., & Brosvic, G. M. (2001). Immediate feedback during academic testing. *Psychological Reports, 88*(3), 889-894.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review, 100*(3), 363.

Fahnoe, C., & Mishra, P. (2013). *Do 21st century learning environments support self-directed learning? Middle school students' response to an intentionally designed learning environment.* Paper presented at the Society for information technology & teacher education international conference.

Fleischhauer, E., Schledjewski, J., & Grosche, M. (2017). Apps zur Förderung von Rechtschreibfähigkeiten im Grundschulalter [Apps to support orthography in primary school age]. *Lernen und Lernstörungen*.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90*(1), 37-55.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88*(1), 3-17.

Google, J. (2019). Android Studio.

Görgen, R., Huemer, S., Schulte-Körne, G., & Moll, K. (2020). Evaluation of a digital game-based reading training for German children with reading disorder. *Computers & Education, 150*, 103834.

Groth, K., Hasko, S., Bruder, J., Kunze, S., & Schulte-Körne, G. (2013). Interventionseffekte bei Lese-Rechtschreibstörung: evaluation von zwei Förderkonzepten unter besonderer Betrachtung methodischer Aspekte [Measurement of treatment in developmental dyslexia under methodological considerations]. *Lernen und Lernstörungen*.

Grotlüschen, A., & Riekmann, W. (2011). leo.– Level-One Studie. *Magazin erwachsenenbildung.at*, 77.

Harrison, R., Flood, D., & Duce, D. (2013). Usability of mobile applications: literature review and rationale for a new usability model. *Journal of Interaction Science, 1*(1), 1-16.

Haßler, B., Major, L., & Hennessy, S. (2016). Tablet use in schools: A critical review of the evidence for learning outcomes. *Journal of Computer Assisted Learning, 32*(2), 139-156.

Heikkilä, R., Aro, M., Närhi, V., Westerholm, J., & Ahonen, T. (2013). Does training in syllable recognition improve reading speed? A computer-based trial with poor readers from second and third grade. *Scientific Studies of reading, 17*(6), 398-414.

Henderson, S., & Yeow, J. (2012). *iPad in education: A case study of iPad adoption and use in a primary school.* Paper presented at the 2012 45th Hawaii International Conference on System Sciences.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. *Psychological Science in the Public Interest, 16*(1), 3-34.

Hochberg, K., Kuhn, J., & Müller, A. (2018). Using smartphones as experimental tools—effects on interest, curiosity, and learning in physics education. *Journal of Science Education and Technology, 27*(5), 385-403.

Huemer, S., Aro, M., Landerl, K., & Lyytinen, H. (2010). Repeated reading of syllables among Finnish-speaking children with poor reading skills. *Scientific Studies of reading, 14*(4), 317-340.

Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E.-M., . . . Valtin, R. (2017). *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich [IGLU 2016: International Comparison of Reading Skills in Primary School]*: Münster; New York: Waxmann.

Klatte, M., Steinbrink, C., Bergström, K., & Lachmann, T. (2013). Phonologische Verarbeitung bei Grundschulkindern mit schwacher Lesefähigkeit [Phonological processing in german poor readers: A pilot study evaluating a computer-based training program]. *Lernen und Lernstörungen*.

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: an 8-year follow-up. *Journal of Educational Psychology, 100*(1), 150.

Logan, J. A., Hart, S. A., Cutting, L., Deater-Deckard, K., Schatschneider, C., & Petrill, S. (2013). Reading development in young children: Genetic and environmental influences. *Child development, 84*(6), 2131-2144.

Mejtoft, T., Lindberg, L., Söderström, U., & Mårell-Olsson, E. (2017). *Feedback in commercial educational applications: Guidelines and conceptual framework.* Paper presented at the Proceedings of the European Conference on Cognitive Ergonomics 2017.

Mkpojiogu, E. O., Hussain, A., & Hassan, F. a. (2018). *A systematic review of usability quality attributes for the evaluation of mobile learning applications for children.* Paper presented at the AIP Conference Proceedings.

Morgan, H. (2013). Multimodal children's e-books help young learners in reading. *Early Childhood Education Journal, 41*(6), 477-483.

Müller, B., Otterbein-Gutsche, G., & Richter, T. (2018). Leseförderung mit Silben und Sprachsystematik: Konzeption eines Trainingsprogramms [Systematic reading improvement using syllables: An intervention program]. *Psychologie in Erziehung und Unterricht, 65*, 52-57.

Müller, B., Richter, T., & Karageorgos, P. (2020). Syllable-based reading improvement: Effects on word reading and reading comprehension in Grade 2. *Learning and Instruction, 66*.

Müller, B., Richter, T., Karageorgos, P., Krawietz, S., & Ennemoser, M. (2017). Effects of a Syllable-Based Reading Intervention in Poor-Reading Fourth Graders. *Frontiers in Psychology, 8*(1635).

Müller, B., Richter, T., & Otterbein-Gutsche, G. (2020). *Lesen mit Willy Wortbär: Ein silbenbasiertes Training zur Förderung der Worterkennung beim Lesen [Reading with Willy Wortbär: A syllable-based intervention to foster visual word recognition]*. Göttingen, Germany: Hogrefe.

Nagler, T., Lindberg, S., & Hasselhorn, M. (2017). Leseentwicklung im Grundschulalter. Kognitive Grundlagen und Risikofaktoren [Reading development in children. Cognitive preconditions and risk factors]. *Lernen und Lernstörungen*.

Naumann, J., Artelt, C., Schneider, W., & Stanat, P. (2010). *Lesekompetenz von PISA 2000 bis PISA 2009*.

Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2018). Educational apps from the Android Google Play for Greek preschoolers: A systematic review. *Computers & Education, 116*, 139-160.

Pfost, M., Hattie, J., Dörfler, T., & Artelt, C. (2014). Individual differences in reading development: A review of 25 years of empirical research on Matthew effects in reading. *Review of educational research, 84*(2), 203-244.

Rashid, T., & Asghar, H. M. (2016). Technology use, self-directed learning, student engagement and academic performance: Examining the interrelations. *Computers in Human Behavior, 63*, 604-612.

Rossing, J. P., Miller, W., Cecil, A. K., & Stamper, S. E. (2012). iLearning: The future of higher education? Student perceptions on learning with mobile tablets. *Journal of the Scholarship of Teaching and Learning, 12*(2), 1-26.

Scherer, R., Tondeur, J., Siddiq, F., & Baran, E. (2018). The importance of attitudes toward technology for pre-service teachers' technological, pedagogical, and content knowledge: Comparing structural equation modeling approaches. *Computers in Human Behavior, 80*, 67-80.

Schoppek, W., & Tulis, M. (2010). Enhancing arithmetic and word-problem solving skills efficiently by individualized computer-assisted practice. *The Journal of Educational Research, 103*(4), 239-252.

Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2015). childLex: A lexical database of German read by children. *Behavior research methods, 47*(4), 1085-1094.

Su, C. H., & Cheng, C. H. (2015). A mobile gamification learning system for improving the learning motivation and achievements. *Journal of Computer Assisted Learning, 31*(3), 268-286.

Sung, Y.-T., Chang, K.-E., & Liu, T.-C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education, 94*, 252-275.

Tacke, G. (2005). Evaluation eines Lesetrainings zur Förderung lese-rechtschreibschwacher Grundschüler der zweiten Klasse [Evaluation of a remedial reading training with dyslexic second graders in primary school]. *Psychologie in Erziehung und Unterricht*(3), 198-209.

Tillmann, A., & Bremer, C. (2017). Einsatz von Tablets in Grundschulen. In J. Bastian & S. Aufenanger (Eds.), *Tablets in Schule und Unterricht: Forschungsmethoden und -perspektiven zum Einsatz digitaler Medien [Tablets in school and teaching: Research methods and perspectives on the use of digital media]* (pp. 241-276). Wiesbaden: Springer Fachmedien Wiesbaden.

Underwood, J., Luckin, R., & Winters, N. (2012). Managing resource ecologies for mobile, personal and collaborative self-directed language learning. *Procedia-Social and Behavioral Sciences, 34*, 226-229.

VERBI Software. (2018). MAXQDA 2018. Berlin, Germany: VERBI Software.

Walter-Laager, C., Brandenberg, K., Tinguely, L., Schwarz, J., Pfiffner, M. R., & Moschner, B. (2017). Media-assisted language learning for young children: Effects of a word-learning app on the vocabulary acquisition of two-year-olds. *British Journal of Educational Technology, 48*(4), 1062-1072.

Maertens, M., Vandewaetere, M., Cornillie, F., & Desmet, P. (2014). From pen-and-paper content to educational math game content for children: A transfer with added difficulty. *International Journal of Child-Computer Interaction, 2*(2), 85-92.

Post-Pandemic Digital Education: Investigating Smart Workspaces within the Higher Education Sector

M A Hannan Bin Azhar
School of Engineering Technology & Design
Canterbury Christ Church University, UK
hannan.azhar@canterbury.ac.uk

Emily Louise Lepore
School of Engineering Technology & Design
Canterbury Christ Church University, UK
emilylouiselepore@gmail.com

Tasmina Islam
Department of Informatics
King's College London, UK
tasmina.islam@kcl.ac.uk

Human collaborations have been challenging with increased dependence on traditional human computer interaction technologies during the pandemic, especially in the education sector. Where conventional learning spaces have not been effective, opportunities have arisen for novel digital innovations to develop a future-proof more realistic learning environment. In the era of post-pandemic digital living, a hybrid approach incorporating Internet of Things and augmented reality will transform future classrooms into highly immersive and collaborative learning spaces. This paper examines smart workspaces within the Higher Education sector and presents an energy efficient, future proof, and collaborative framework for smart learning environments.

Smart workspace, Higher education, Post-pandemic, Augmented reality, Internet of things.

1. INTRODUCTION

During the time of global pandemic, everyone has had to adapt to remote working and remote learning, especially staff and students at higher education institutes have adapted to a new way of working, either remotely or in hybrid mode based on the severity of the state of the pandemic (Baker and Abbo, 2020). Reliance on technology-based solutions to enable education has been vital during this challenging time. 'Smart Education' refers to learning outside a traditional classroom; and is an activity that can be done anywhere and anytime (Bajaj et al., 2018). Abdel-Basset et al., (2018) suggested that use of IoT devices within the higher educational sector could inspire smarter lesson plans, enhance access to information and design of more secure campuses. By having smart campuses, it means smart timetabling, parking, energy management and smart learning can all be managed, so students can have improved and elevated experiences (Afshar et al., 2020). Better system management leads to considerable reductions in energy consumption, reducing environmental impact through more efficient lighting, thermal and ventilation control.

A collaborative workspace is all about bringing a team together to pursue a common goal (SpaceIQ, 2020) whereas individual workspaces are to do with the space required to complete one's work by themselves. Companies are now embracing innovative space designs like studios, pop-ups and hotdesking instead of traditional offices as these provide good mobility and flow of shared spaces (Khan, 2019). Like the workplaces, smart workspaces in the education sector will allow students to have an enhanced and interactive learning experience, save time and of course, increase productivity (Kaur, 2018). Smartboards and virtual reality are also extremely important in both office and university spaces. Individuals are able to participate in demos as if the product was right in front of them. Augmented reality (AR) should enhance the experience in a learning space as it adds digital elements onto a smart screen (such as Microsoft Hololens, smartphone screen etc.) creating an illusion that holographic content is within the physical world around you (AVRSpot, 2018). AR can provide a view of the real world by using computer-generated information and offers a convenient and intuitive way for users to visualize and interact with IoT objects and their data (Dongsik et al., 2019). Current literature (Wired, 2018; Khan, 2019) discusses the need for smart technologies for generic office spaces without detailing how they can be adapted for learning in an education sector. With a preliminary design specification, this paper details the architectural framework for an IoT based smart and energy efficient workspace enabling interactive and immersive learning experiences for education. In order to justify design choices for smart workspaces in education, a survey was conducted to gather student opinions. Also, examples of existing smart office spaces from literature (e.g. Wired, 2018;Gracia, 2017 etc.) will be given to support arguments. Students currently enrolled in UK universities or those who recently graduated from one were included in the survey. From ten different universities in the UK, 44 responses were received to assess student experiences with higher education technologies. 47.7% of the participants were male, 52.3% were female, and they had varying levels of technological proficiency and came from different disciplines (both science and humanities); few had specific learning needs. Participants were asked to rate their opinions on a Likert scale on questions related to the impact smart learning spaces might have on University education, e.g. adaptation of ergonomic factors and automatic

climate control around study spaces, flexible learning experiences including interactive remote collaboration and immersive AR based face-to-face learning, privacy issues related to GPS location tracking and health data monitoring through mobile Apps etc. Ethical approval was obtained, and participants gave consent to take part in the study.

The remainder of this paper is organized as follows: Section 2 describes the architectural layers of the proposed framework, Section 3 details the key components, example layouts and benefits of smart workspaces in the higher education sector and finally, Section 4 concludes the paper.

2. PROPOSED FRAMEWORK

The following section describes the design framework for a smart workplace within a university setting. The framework is composed of three layers: Internet of Things layer, Edge layer, and Education Solution (Figure 1), which are described below.

2.1. IoT layer

The Internet of Things (IoT) layer includes devices such as sensors, actuators and controllers that enable monitoring of equipment in operation and human activity. The layer uses a wireless standard such as ZigBee or Wi-Fi. ZigBee can be more appropriate for constrained IoT devices due to its low power consumption and long battery life (ZigBee, 2021). The devices in this layer produce low level time series data such as values of temperature, humidity, heart rate, movement, solar radiation etc. Devices that use the Internet of Things facilitate significant energy savings from efficient usage of lighting, heating, air conditioning, etc., as well as the locating of staff/group members by a real-time locating system. To optimise the efficiency, dimension reduction will be used on the raw data by finding repeated patterns (Muhammad, 2020) before sending it to the edge layer. Data reduction can prevent I/O bottlenecks, as well as dramatically reduce storage, bandwidth, and energy costs.

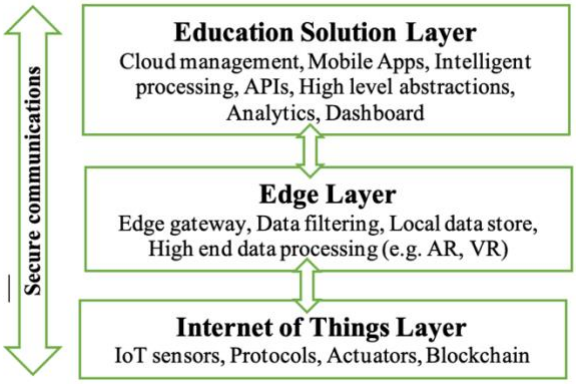


Figure 1: *Architecture of smart learning Space*

2.2. Edge layer

The next layer is the Edge layer, which is the central layer that monitors and updates the technological resources needed to manage the students' activities. The data is filtered in this layer before its being sent to the cloud to be used by the educational applications. Microcomputers such as Raspberry Pis with additional storage capability could be used for low-cost solutions at this layer. In scenarios like VR and AR applications, faster high spec computers and routers can serve as edge gateways for high end data processing and storage. It is possible to use different communications standards depending on the IoT scenarios and uses, such as Wi-Fi, Bluetooth, ZigBee, mobile telephony, etc.

2.3. Education solution layer

Lastly, there is the education solution layer which is a set of services and Artificial Intelligence (AI) assisted applications to develop intelligent interaction interfaces for development within the education sector. In this layer, the use of APIs allows smart cloud services to be available through web browsers or mobile apps. The aim is to provide edge node management and security, improving energy efficiency within the education workspace. For example, the use of a mobile app to locate or track down staff/group members to increase productivity would be extremely efficient. Apps should also ensure that students are aware of saving energy and changing their behaviours so the resources can be used efficiently.

2.4. Secure framework

Data integrity and reliability will be ensured using blockchain IoT paradigms (Reyna et al., 2018). For example, use of automatically executed smart contracts at the IoT nodes can ensures that no data are modified or corrupted by any other element of the network. The proposed distributed architecture for data storage and processing at the edge layer will have an advantage over centralised cloud-based

solutions (Manna et al., 2021) as the latter could be vulnerable to denial-of-service (DoS) attacks, where attackers could flood the system and attempt to shut down the server. With the proposed secure architecture, DoS attacks will be prevented, ensuring better response times for systems. Due to the sensitive nature of the data (e.g. student's health data, student id etc.), while data are processed in transit (e.g. over the physical and network layer before reaching the cloud), unsecure communication can be susceptible to intrusions such as the man-in-the-middle attack; it is important that all data transmission should include strong secure encryption techniques and an automated intrusion prevention system should be in place to detect unwanted intrusions and prevent malicious activities.

3. KEY COMPONENTS AND BENEFITS

Figure 2 illustrates the key components required to detect and track human presence in a learning space and to measure light levels, temperature, humidity, air quality, etc., to manage energy costs effectively. The example rooms (Figures 3 and 4) incorporate the key components and IoT-based Edge Computing services in order to provide faster, more efficient, and secure responses beneficial to student learning. Smart cameras in both rooms can track what is happening by allowing video transmissions for others to view especially those working from home. In both rooms, cloaking the windows is important since it provides transparency and security. As a result, screens cannot be viewed from the outside, which is ideal when confidential information is displayed. To ensure maximum participation, both remote and face-to-face collaborations are incorporated into the design. There are spaces in the room for AR based 3D projections. Students can use smart AR glasses for immersive practical experience when the real work is happening in a remote location. As an example, a virtual demonstration of a human body in a hospital will be transferred to the classroom and medical students wearing AR glasses will be able to see and interact with the demonstration in real time. Further benefits are discussed below.

3.1. Energy efficiency and sustainability

The chairs are designed with sensors that adjust automatically so students of different heights can sit ergonomically. In the survey, when students were asked about automatic adaption of ergonomics and the adjustability of the workspace, nearly 90% said this was beneficial. As standing desks have been scientifically proven to improve productivity, concentration and creativity (Posture Group, 2019), they can be used for both collaborative and individual workspaces. After a certain period of time, smart apps could alert students and staff to stand up, and then the desk will become a standing desk.

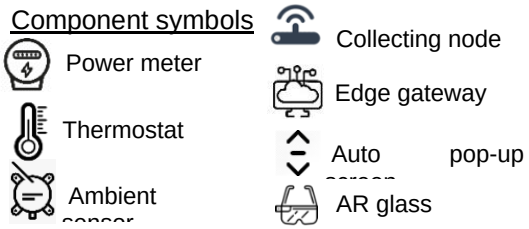


Figure 2: *Symbols of key components*

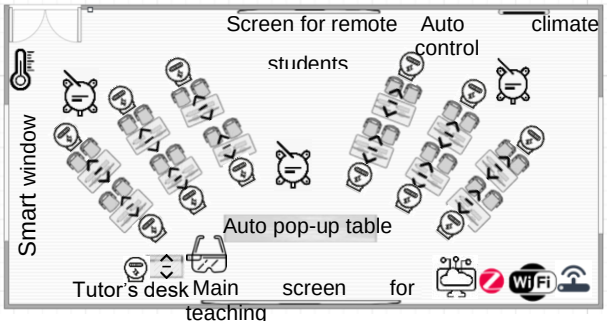


Figure 3*: Smart classroom design*

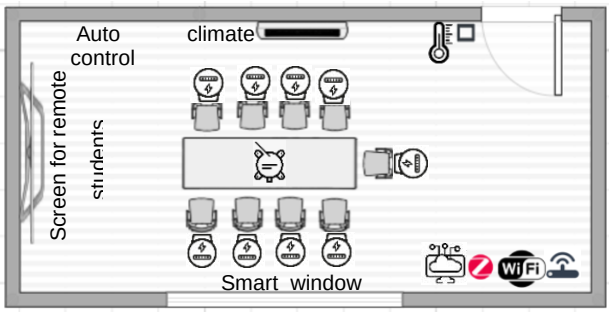


Figure 4: *Smart meeting room design*

In the survey, when students were asked about automatic climate control, more than 80% responded that changing the climate depending on their needs would help them focus on studying and increase their productivity. This led to sensors being included in the building design so that temperature, lighting, air quality etc. can be adapted in order for students to get the most out of their studies (Azhar et al., 2017). Edge gateways and collecting nodes will gather information from the workspace, such as temperature, seat settings, fitness levels, etc., to analyse further and make autonomous control of those factors to improve performance and satisfaction during learning. The windows in the rooms (Figures 3 and 4) are smart and regulated automatically to suit best lighting conditions for the intended work. Furthermore, solar panels for water heating and photovoltaic panels on the roof will reduce carbon dioxide emissions significantly. The Edge in Amsterdam, for instance, uses 70% less power than comparable buildings (Unique Tech News, 2017). A reward point system can be used to encourage energy saving habits in students, e.g. obtaining daily power consumption below the

previous day's average, taking the stairs instead of the elevator, etc. By collecting enough points, students can get free juice or drinks in the cafeteria.

3.2. Smart inventory management

With a mobile app, users can book a meeting room, as shown in Figure 4, almost instantly. Students can book the room immediately after it becomes available online, since availability can be updated in real time. In the survey, when students were asked about the automatic allocation of study rooms based on the schedules, preferences and availability, 82% agreed that this will increase their productivity by saving time through a transparent autonomous inventory management system. The same concept is demonstrated in an office building in Paris (Majunga Tower, 2021) that allows users to choose a workspace, book a meeting room, control their environment via tablet or phone to support wellbeing and productivity. Also, the Edge in Amsterdam has an IoT infrastructure for inventory management, which has significantly increased employee satisfaction and reduced absenteeism by 60% (British Land, 2017). Furthermore, smart classrooms will be equipped with appropriate sensors to detect component faults (e.g. network printers, routers, etc.) and with AI-assisted software, it will be able to predict system faults and, thus, prevent system downtime. An Italian hospital, for example, used smart fault control to predict 61% of real faults, including 93% of overheated appliances (British Land, 2017).

3.3. Smart wellbeing

Using wearable sensors in smart gadgets (e.g. smartwatches, mobile phones), it is possible to measure a student's stress level or heart rate and flag up concerns when the condition crosses a threshold. If there is a developing health risk, such as a panic attack, university wellbeing teams can be contacted automatically via text messages. In addition, indoor air quality, light, and noise levels can negatively impact performance and wellbeing. In order to deal with mental health issues, such as depression and dissociation at work, changing colours, circadian lighting systems, and arranging the study spaces according to the time of day can be beneficial.

3.4. Smart social networking

Scientists at MIT (British Land, 2017) have developed social physics, which uses data from phones, wearables, and sensors to shed light on how people communicate. An employee's wearable badge can record their whereabouts and who they meet, as well as the nature of the discussions. Students can also take advantage of smart campuses by ordering lunch, booking a meeting room, or tracking staff or other students using mobile apps. Tracking in conjunction with the campus map can serve as a visual guide to find someone. However, despite the fact that tracking a person or measuring health data for wellbeing might be beneficial, survey results show that the majority of students (79.5%) were not in favour of personal data collection by smart educational institutions. Student perceptions indicated that personal data collected (e.g. tracking GPS locations, health data from wearable devices, etc.) could easily be misused, leading to more harm than good.

3.5. Inclusive system

The survey showed that participants with learning disabilities requested accessibility tools to assist them in learning. In order to ensure that all students have access to the content by default, these resources should be included at the interface level of the smart space. An assisted mobile app can provide resources such as screen readers, tinted screens, options for larger text, text-to-speech software, and blue visors for reading and taking notes. In order to increase the satisfaction level, the default values of the app can be customised to meet the needs of each learner. The app can change the default settings to custom values to the satisfaction level of the learners. With the assistance of AI, app can recommend the optimal settings for each learner.

4. CONCLUSION

In this paper, we proposed architectural layers which can be used to implement an IoT-based learning space. The survey findings have helped get a better understanding of what students want and how their viewpoints can influence the design of the new smart workspace for learning. The paper described the three-layered architecture of the framework. With the proposed architecture, the computation load is balanced between the cloud and the edge, which results in a reduction of computational capacity, consumption, and hardware resources at the physical or IoT layer. The education solution layer on the cloud allows AI-assisted applications to identify the optimal settings for information fusion and training. Smarter IoT controlled efficient lighting, thermal and ventilation will lead to considerable reductions in energy consumption to reduce environmental impact meeting low-carbon targets set by the government. In cases where privacy concerns are raised in the survey, the framework justifies the security of the data by proposing the use of the blockchain at the edge and strong encryption. Specifically designed for security and energy efficiency, the framework offers a sustainable learning environment for higher education in the new post-pandemic digital age.

5. REFERENCES

Abdel-Basset, M., Gunasekaran, M., Mohamed, M., and Rushdy, E. (2018) Internet of things in smart educational environment: Supportive framework in the decision-making process, Wiley, Special issue paper, Pages: 1-12.

Afshar Alam, M and Saiyeda, A. (2020) A Cloud Based Solution for Smart Education. Volume 11, Issue 2. Pages 28-29.

AVRSpot (2018) What is Augmented Reality (AR) and how does it work? Available at: https://www.avrspot.com/augmented-reality-ar-work/ (Accessed: 28/06/2021).

Azhar, M.H.B., Islam, T. and Alfieri, M. (2017) Effects of students' preferences in use of lighting and temperature on productivity in a university setting. In *EAI International Conference on Technology, Innovation, Entrepreneurship and Education* (pp. 139-145). Springer, Cham.

Baker, M. D. and Abbo, T., (2020) Working Post-Pandemic: What Campus Employees Need (Now), Available at: https://er.educause.edu/blogs/2020/5/working-post-pandemic-what-campus-employees-need-now (Accessed: 28/06/2021).

Bajaj, R and Sharma, V. (2018) Smart Education with artificial intelligence-based determination of learning styles, International Conference on Computational Intelligence and Data Science, ICCIDS, Pages 835-841.

British Land (2017) Smart offices A 2017 vision for the future, Available at: https://officeagenda.britishland.com/assets/pdfs/smart-offices.pdf (Accessed: 28/06/2021).

Dongsik, J. and Gerard, K. (2019) AR Enabled IoT for a Smart and Interactive Environment: A Survey and Future Directions, Sensors, PMC.

García, Ó., Prieto, J., Alonso, R. and Corchado, J. (2017) A framework to improve energy efficient behaviour at home through activity and context monitoring. Sensors, 17, 1749.

Kaur, D. (2018) How Smart Class Technology is Benefiting Education Sector. Available at: https://www.entrepreneur.com/article/322587 (Accessed: 28/06/2021).

Khan, B. (2019) Trends in Workspace Design with IoT and Tech. Available at: https://www.wrld3d.com/blog/trends-in-workspace-design-with-iot-and-tech/ (Accessed: 28/06/2021).

Manna, S.K., Azhar, M.A.H.B and Sakel, M. (2021) A Smart and Secure IoMT Tele-Neurorehabilitation Framework for Post-stroke Patients, International Conference on Industrial Instrumentation & Control (ICI2C-2021), to be appeared on Lecture Notes in Electrical Engineering, Springer ISSN: 1876-1100.

Majunga Tower (2021) Eiffage construction's showcase Project of Majunga Tower in Paris, http://www.eiffageconstruction.com/expertise/showcase-projects/majunga-tower (Accessed: 28/06/2021).

Muhammad, F.M.M. (2020) Extreme-SAX: Extreme Points Based Symbolic Representation for Time Series Classification. In: Song M., Song IY., Kotsis G., Tjoa A.M., Khalil I. (eds) Big Data Analytics and Knowledge Discovery. DaWaK Lecture Notes in Computer Science, vol 12393. Springer, Cham. https://doi.org/10.1007/978-3-030-59065-9_10

Posture Group (2019) Benefits of sit stand desks on your posture and work health & well-being. Available at: https://www.posturegroup.co.uk/blog/benefits-of-sit-stand-desks-on-your-posture-and-work-health-and-wellbeing (Accessed: 28/06/2021).

Reyna, A., Martín, C., Chen, J., Soler, E., Díaz, M. (2018) On blockchain and its integration with IoT. Challenges and opportunities. Future Gener. Comput. Syst., 88, 173–190.

SpaceIQ (2020) What is a Collaborative Workspace, really? Available at: https://spaceiq.com/blog/what-is-a-collaborative-workspace/ (Accessed: 28/06/2021).

Unique Tech News (2017) Smart office solutions in the world's most sustainable building, Available at: https://uniquetechsolutions.co.uk/smart-office-solutions-in-the-edge-amsterdam/ (Accessed: 28/06/2021).

Wired (2018) The Workplace of the Future. Available at: https://www.wired.com/brandlab/2018/06/the-workplace-of-the-future/ (Accessed: 28/06/2021).

ZigBee (2021) Why Zigbee? Why should I use Zigbee in my products and services? Available at: https://zigbeealliance.org/why-zigbee/ (Accessed: 28/06/2021).

The Impact of Ramadan on Online Learners Behaviour: What Can We Learn from Tracking Learners' Data?

Wad Ghaban
School of computer science
University of Birmingham
B15 2TT
whq360@alumni.bham.ac.uk

Robert Hendley
School of computer science
University of Birmingham
B15 2TT
r.j.hendley@cs.bham.ac.uk

Rowanne Fleck
School of computer science
University of Birmingham
B15 2TT
R.fleck@cs.bham.ac.uk

The flexibility of online courses means that people are able to fit them around other commitments, often studying duringholiday times. However, there is limited research examining the behaviour of online learners during holidays and specialevents. It is important to understand this in order to know when learners will be active and therefore when they might becontactable or require online support. We conducted this study to examine the changes in behaviour of online learners during Ramadan, a period when Muslims are required to fast between sunrise and sunset. This change in mealtimes affects Muslims' wake-sleep cycles and work patterns as well as all other aspects of their lives. The results can be used toinfer an understanding of their lifestyle during this period.

The insights that this provides are important in the context of online learning but also many other application domains where users' availability and openness to interaction are important.

Tracking users, Online learning, Behaviour, Holiday, Ramadan, Activity

INTRODUCTION

Online learning is popular because of its flexibility - learners can learn at the time and location of their choice. This flexibility attracts many users during the holidays. However, schedules and sleep patterns can be different when students are not in school/college or at work. These changes in behaviour should be accounted for when designing online courses. For this kind of courses, it is important to know when learners are available and able to interact with live lectures (Shelton et al. 2017). There is little research exploring how learners interact with online courses during special occasions such as holidays. This may be because of the difficulty of reaching participants during these times (Harrison et al. 2015); (Vlajic et al. 2018). One of the special holidays that has been widely accepted as a subject of interest is the holy month of Ramadan. Ramadan is a religious month for Muslims. All healthy, adult Muslims are required to refrain from eating any food, drinking any liquid or taking any oral medication during daylight hours (Ramadan 2002). Most of the research studies addressing this period have investigated the effect of fasting on Muslims' body fat and their health. A few studies have tried to examine the behaviour of Muslims during Ramadan (Roky et al. 2001); (Reilly and Waterhouse 2007).

However, the results from these studies may not be reliable due to issues related to their methodology, particularly the small number of Muslims involved in the studies. Further, these studies were based on self-report questionnaires.

In the light of this gap in the research, we ran a study using a more objective approach by tracking the activity of 564 learners to observe their usage patterns when using an online learning website.

The results from this study, and similar studies, show that during Ramadan, the optimal time to contact learners is during the night. These results provide important data that can be used to build hypotheses about the behaviour of users of online courses. A central goal of this research is to inform the design of personalised models that can be used to drive adaption to match both individual users' characteristics and variations in behaviour over time, especially during special events.

This research is important for applications such as online learning, but also for other domains where understanding users' behaviour and when to interact with users is crucial. For instance, adaptive notification systems need to build models to predict opportune moments for interruptions. While these models often make allowances for different behaviour during weekends, they should also be able to cope with special periods such as religious festivals, celebrations, illness and so on.

BACKGROUND

Learning is a continuous and active process (Ally 2004). However, researchers have found that over holidays, students lose some of what they have learned in school (Shelton et al. 2017). To mitigate this drop, many families encourage their children to take advantage of online courses during the holidays (Shinwell and Defeyter 2017). To make these online courses more effective, it is important to understand how user behaviour changes and when students are available. Knowing when users are accessing the platform can help the designers to serve users better and provide them with a better user experience.

One popular holiday in Muslim community is Ramadan in which the altered cycle of sleeping and waking is changed (Roky et al. 2001). Several studies have sought to document changes in sleep-wake patterns during Ramadan. For example, Wilson et al. (2009) examined 20 male soccer players to explore whether there was any significant difference between their sleep-wake patterns during Ramadan compared to normal times. The results showed that their sleeping and waking times during Ramadan differed significantly from their sleeping and waking times in other months. These results support the finding from (Roky et al. 2001). In this study, they asked 8 Muslims to report their sleep time and sleep duration before, during and after Ramadan. They found that sleep time is delayed and sleep duration is decreased during Ramadan.

Most of the previous studies have been based on self-reported information that was obtained by asking participants about their behaviour during Ramadan (Mehmoodet al. 2015); (Yıldırım-Yenier et al. 2016). They also have quite small numbers of participants. Further, to the best of our knowledge, no study has tried to understand how individuals learn or study during Ramadan. This is partly because Ramadan is a public holiday in most Muslim countries. However, in some countries, many people work during Ramadan. Even on working days, however, Bahammam (2006) shows that these Muslims follow a specific pattern by staying up all night and resuming their sleep after finishing their work.

To address this lack of understanding, we designed the current study. The aim is to analyse the behaviour of online learners during Ramadan by monitoring their use of an online learning website and comparing this with usage patterns outside of Ramadan.

Method

This study aimed to develop an understanding of the waysin which online learners interact with a learning website during Ramadan, and what subtle changes occur in their behaviours.

For this purpose, we asked 564 Muslims learners in Saudi Arabia to register in an online learning platform. After registration, participants were divided in two groups: one group, 194 Muslims learners (91 boys, 103 girls), used the online learning website before Ramadan (4th–29th September 2018), and the second group, 370 Muslims learners (169 boys and 201 girls), used the same online learning website during Ramadan (10th May–4th June 2019). During the designated time period, learners were free to use the website at any time and in any place. Further, we told the learners they were free to drop out of the study whenever they wanted. The platform supported with chat feature, which participants are able to talk to their friends by clicking on a button called 'talk to a friend'.

In the beginning of our study, we measured the number of learners who interacted and engaged with the online learning website during Ramadan and non-Ramadan days. We calculated the number of active learners as each time the learner accessed the website; if a learner accessed the website twice in a single hour, the system registered the activity two separate times.

In 2019, Ramadan began on Monday, May 6th. and continued until Monday, June 3rd.. During this time, Muslims are expected to fast from sunrise until sunset. The time of day at which the sun rises and sets varies according to geographical location and changes throughout the period. We collected all the data in this study from participants in northwest Saudi Arabia. The times at which the sun rises and sets change few minutes between the beginning and the end of Ramadan.

The system began recording users' activities during Ramadan on May 10, 2019, which was the fifth day of Ramadan. Because we were only concerned about the number of users who were active during this time, the system only registered the number of users accessing the website. We used this dataset to determine whether there was a common pattern in user activity. We then compared this data with data from the non-Ramadan days.

RESULTS

Figure 1 shows the average percentage of learners who were active online on both the Ramadan and non-Ramadan days. During Ramadan, learners began their activity after breaking their fast (after sunset). Further, these learners stayed awake all night and into the early hours of the morning. This contrasts with the activity of the learners on normal (non-Ramadan) days, when they were more active and interacted with the Internet during the day, especially after school.

Because of the interesting change in the learners' behaviours during Ramadan, we examined the data we obtained during Ramadan in more detail. We monitored the number of active daily learners from May 10 to June 4, the last day of Ramadan. We notice some dips in the usage of the online learning system on specific days, such as the 17th, 24th and 31st of Ramadan. We found that all of these dates

were Fridays.
Further, the number of active learners rises until the middle of Ramadan before dropping at the end of Ramadan.
After plotting the total activity of all the learners, we examined whether there is a variation between the activity of the boys and the girls (figures 2,3,4 and 5.)

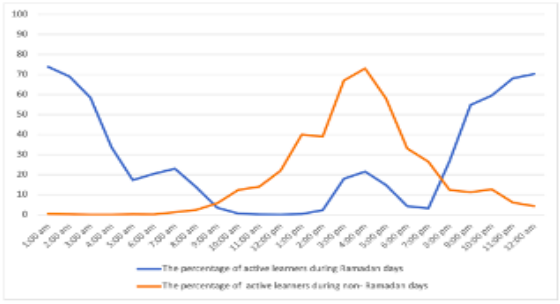


Figure 1: *The variation in the activity of the users during both Ramadan and non-Ramadan days*

DISCUSSION

The results from our study supported the findings from related research studies (Roky et al. 2001); (Bahammam2006); (Reilly and Waterhouse 2007). Most Muslims became active during the night, which is considered an interesting phenomenon and is not part of any cultural tradition. Most Muslims interacted with the online learning platform during the night which is the standard time to eat and drink. In addition, this study took place in Saudi Arabia, where this Ramadan occurred in the summer season. The accompanying high temperatures may have had an effect on user behaviour, as users may have preferred to stay home and sleep rather than be awake and active in the hot weather during the day. This became clear when we analysed the messages exchanged between learners. For example, one learner typed, 'Why should I wake up early, there is nothing to do and it is hot!' Another learner wrote that he or she slept during the day because most places, such as shops and malls, were closed. As one learner wrote, 'If I wake up earlier, what I should do. All [the] malls and shopping centres are closed, and of course, restaurants are closed as well.'

Another interesting point to consider is the observed variation in behaviour between boys and girls. The overall numbers of boys and girls who were active on the website were quite similar during each hour over the entire month. However, we noticed some differences in their behaviour. For example, in figures 2,3,4 and 5. we see that girls were active earlier in the day than boys, and that boys stayed awake later. Further, the number of active learners changed over Ramdan. This number rose by the middle of the Holy Month.
Finally, there were again only a few active learners during the days before the end of Ramadan, with a notably smaller numbers of girls. This can be explained by preparations for Eid. In most Muslim countries, it is traditional that women spend time shopping for new clothes, furniture and decorations around this time to celebrate the end of the fast.
Another observation made during this study is the variation in learners' behaviour between different days of the week. For example, Friday is considered a special day of the week, as male Muslims must attend mosques to pray a special prayer in the afternoon called aljuma'a. Accordingly, we noted that more learners were active in the late morning on Fridays.
Figure 6 shows learners' activity on Fridays and compares it to the average learners' activity on other days of the week.

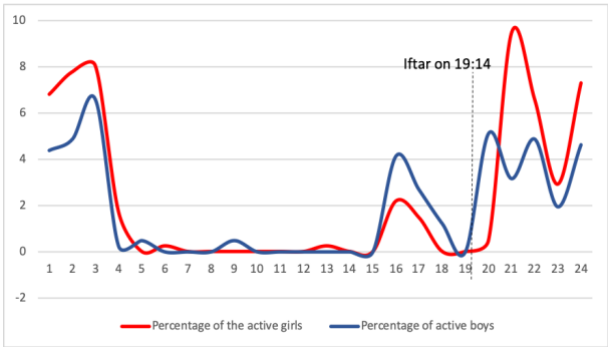


Figure 2: *The activity of users on the 11th. May*

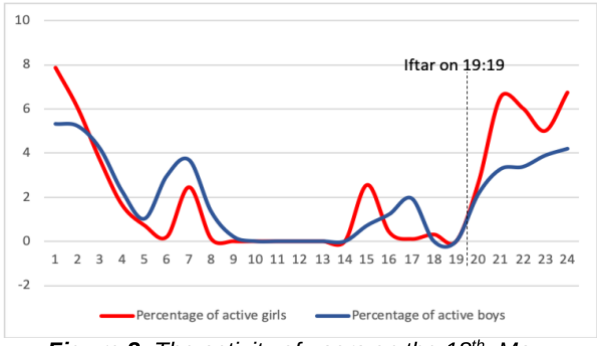


Figure 3: *The activity of users on the 18th. May*

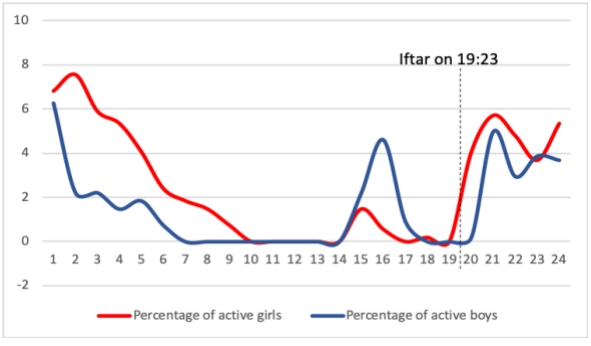


Figure 4: *The activity of users on the 25th. May*

Figure 5: *The activity of users on the 1st. June*

Our results could help identify the optimal time to contact Muslims during Ramadan, in particular in relation to when to publish online content or for advertising. Finally, the findings could also be compared to the results of other studies to identify similarities and differences between the behaviour of Muslims living in different regions. This data can be used as part of digital ethnography research to develop insights into Internet-user behaviour during Ramadan.

In this research, we collected data from teenage learners who voluntarily visited a learning website, but the results may be generalisable to other Muslims living in the same geographical area (Saudi Arabia). Our results confirmed what what Al-Ghadir and Azmi (2019) suggested. In their research, they tracked the number of tweets from adults during Ramadan.

Of course, this study was undertaken with a very narrow user group: young adults in one geographical area. Further work should be done to explore whether these results can be generalised to other age groups and to other countries. In order to obtain more generalisable data, we may need to repeat the study with predominantly Muslim users in other geographical locations. However, Ramadan occurs only once a year, which makes it difficult to repeat the study multiple times. In addition, we only focused on teenage learners, who do not have as many responsibilities as adults. Generalising the results to other populations would require tracking the activity of adults using online learning courses or other online systems.

CONCLUSION

The aim of this research was to investigate the behaviour of online learners in Saudi Arabia during Ramadan. Within this work we argue that this change in the observable on-line behaviour can be used to provide insight into the broader changes to users' pattern of living. The results from this study show that there is a variation in the active time of the learners during Ramadan versus non-Ramadan days.

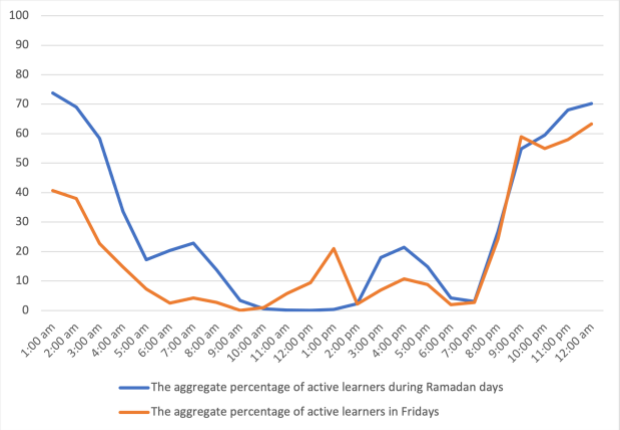


Figure 6: *The activity of the users in Fridays compared with theother days*

Our results provide an understanding of learners' behaviour during Ramadan. This understanding is important in the context of online learning. For example, it may possible to use the results from these studies to build an adaptive model that presents the learning content based on learners' optimal time. Further, the results also can be used to identify the best time to connect with Muslims during Ramadan.

The insights provided by the analysis of this digital data are important. Although these changes in behaviour have been understood anecdotally, this is the first study to report results based on objective data and with a significant number of research subjects.

This kind of study is important for investigating the behaviour of Muslims during Ramadan, and similar approaches could be used to explore other special events and periods. While many adaptive systems do take account of the differences between weekdays and weekends, there are actually a very large number of exceptional periods; thus building an average model of behaviour that ignores these differences will only have limited success.

There are many applications of our findings to domains in which a model of user availability might be used to determine the most appropriate time to issue notifications or interact with users. For instance, choosing the right time to push a social media notification or to post an advertisement has a large impact on its effectiveness.

The results from the study reveal the difficulty of buildinga constant model for users during different times and occasions. However, it is important to improve the user experience by building a dynamic model that recognises the time and the user status, then provides customised content based on the provided information.

REFERENCES

Al-Ghadir, A. I. and Azmi, A. M. (2019), 'A study of arabic social media users—posting behavior and author's gender prediction', Cognitive Computation 11(1), 71–86.

Ally, M. (2004), ‘Foundations of educational theory for online learning’, Theory and practice of online learning 2, 15–44.

Bahammam, A. (2006), ‘Does ramadan fasting affect sleep?’, International journal of clinical practice 60(12), 1631–1637.

Harrison, D., Marshall, P., Bianchi-Berthouze, N. and Bird, J. (2015), Activity tracking: barriers, workarounds and customisation, in ‘Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing’, ACM, pp. 617–621.

Ramadan, J. (2002), ‘Does fasting during ramadan alter body composition, blood constituents and physical performance?’, Medical Principles and Practice 11(2), 41–46.

Reilly, T. and Waterhouse, J. (2007), ‘Altered sleep– wake cycles and food intake: The ramadan model’, Physiology & behavior 90(2-3), 219–228.

Roky, R., Chapotot, F., Hakkou, F., Benchekroun, M. T. and Buguet, A. (2001), ‘Sleep during ramadan intermit- tent fasting’, Journal of sleep research 10(4), 319–327.

Shelton, B. E., Hung, J.-L. and Lowenthal, P. R. (2017), ‘Predicting student success by modeling student interaction in asynchronous online courses’, Distance Education 38(1), 59–69.

Shinwell, J. and Defeyter, M. A. (2017), ‘investigation of summer learning loss in the uk-implications for holiday club provision’, Frontiers in public health 5, 270.

Vlajic, N., El Masri, M., Riva, G. M., Barry, M. and Doran, D. (2018), Online tracking of kids and teens by means of invisible images: Coppa vs. gdpr, in ‘Proceedings of the 2nd International Workshop on Multimedia Privacy and Security’, ACM, pp. 96–103.

Wilson, D., Drust, B. and Reilly, T. (2009), ‘Is diurnal lifestyle altered during ramadan in professional muslim athletes?’, Biological Rhythm Research 40(5), 385–397.

Can Educational Software Support Learning in the Global North and South Equally? A Comparison Case Study

Pamela Andrade
University of Leicester
Leicester - United Kingdom
pyas2@le.ac.uk

Effie L-C. Law
University of Leicester
Leicester - United Kingdom
lcl9@le.ac.uk

Nowadays, many educational software applications are designed with input only from specific user groups from developed countries in the Global North, without considering the needs of users living in the Global South, potentially compromising their effectiveness. To contribute to the understanding of how educational software influences learning depending on the country of origin of end-users, this paper reported results of learning performance and user experience (UX) from 176 students from Ecuador in the Global South, and compared those from 196 students in the United Kingdom in the Global North. Results in the Ecuadorian sample showed that there were some significant differences on the students' perception of software usability and their self-reported measure of user engagement, but no significant differences were found on their learning gains unlike the sample of students from the UK. Comparisons between the schools in the two countries showed some differences in the students' attitudes and motivations towards learning science prior to the study. We discussed how these could have influenced students' learning performance and UX.

Technology-enhanced learning, Gamification, User Experience, Digital Divide.

1. INTRODUCTION

An exponential growth of computing devices and a widespread access to the Internet have changed over the past few decades the ways in which people communicate, work, and learn. In fact, we have seen and experienced over the last couple of years a drastic change in typical human habits of software use due to the Covid-19 pandemic, and we have witnessed a visibly larger scale of acceptance and adoption of online educational software, ascertaining its utility and necessity in the modern world. Concomitantly, the Human-Computer Interaction (HCI) community has been exploring how modern technologies shape the development of future education in the quest of understanding how software applications can influence people's routines and behaviours, and vice versa [1]. This paper focuses on how educational methods have adapted to the generalised use of the Internet, and how educational software influences learning on students in the Global North (high-income countries) and the Global South (low-income countries).

Technologically speaking, ongoing debates in the HCI community have brought up the issue that, far from being universal, software design practices are still inherently based on westernized models that are applied in the Global South without accounting for local knowledge systems that could impact their interactions with the software [2]. Learning environments that are aware of social and political dimensions are therefore critical in the development of better technology that supports universal learning and knowledge sharing [3]. In an attempt to contribute to the understanding of how educational software influences learning in the context of the digital divide. this paper mainly collects responses from students in two secondary schools in a Global South country (Ecuador), but it also contrasts these results with those collected (see [4]) from two secondary schools participating in a similar study in a Global North country (The United Kingdom).

In addition, this study investigates how the inclusion of a popular educational theory, namely *Learning by* Questioning, and an engagement strategy, namely *Gamification*, could improve performance in young students and enhance their user experience. Hence, the contribution of this research is twofold: (i) comparing individual user interface (UI) design elements in an educational application with the objective to analyse which specific *gamification elements* are more beneficial to students; and, (ii) to contrast the responses of students in two different countries (one in the Global North and one in the Global South) to analyse if our *gamified questioning-based* approach positively influences students regardless of their country of origin. This work is framed within the Go-Lab platform[3], which provides online tools (such as lesson content and learning apps) to teach STEM subjects to students in different age groups and in several languages. The Go-Lab platform provided the context of the activity, where all students learnt about the same lesson topic, created questions about it, and earned a reward in the form of a digital gamification element for their participation (see Section 3.6). Learning apps were designed and developed for our studies with the input from teachers and students in Ecuador and the United Kingdom. For the purpose of rewarding students (see Section 3.5.3) gamification elements were embedded in *Graasp*[4] to analyse with qualitative and quantitative data collection methods (Section 3.4) how these influence UX in terms of perceptions of software usability, user engagement, and motivation, as well as student performance.

2. CONTEXTUAL BACKGROUND

This section starts by addressing the topic of educational technologies, followed by the discussion how gamified systems and the most salient gamification elements in the related literature have been implemented and studied. This section also explores how students could benefit from a popular educational theory, which is the pedagogical backbone of this research. At last, this section discusses the Global North-South divide.

2.1 Internet Access and Digital Educational Technology

Nowadays, most people connect and interact with each other via the Internet, making it an indispensable collaborative tool [5]. In 2006, 57% of households in the United Kingdom had access to the Internet; and this number elevated to an astonishing 96% in 2020 [6]. Although this broader availability of online solutions is really a global trend, countries in the Global South have shown a slower growth of households having access to the Internet. In the particular case of Ecuador, 22.5% of households had Internet access in 2012, increasing to only 45.5% in 2019 [7]. Potentially, this could be problematic for low-income countries and students that have been forced to stay at home during the pandemic. Nonetheless, as the access to reliable educational materials is no longer restricted to traditional lectures, students and teachers have been adapting their routines to utilise the benefits of modern technology.

Likewise, with the generalised use of the Internet, schools and universities have similarly seen an increase in the number of technological tools available to their students, in and outside the classroom. However, this availability of educational resources online does not guarantee their favoured use, especially when employed as extracurricular activities for young students to do on their own, as many of them would still prefer to spend their time on the Internet on

[3] *www.golabz.eu* is the official Go-Lab *sharing* website, generally used to distribute online lessons and showcase different apps and labs.

[4] *www.graasp.eu* is the Go-Lab *authoring* website, used to create learning spaces and support interactive lessons with different apps and labs.

more recreational and ludic ventures, such as playing games [8]. Consequently, a number of approaches and techniques have increasingly been utilized in educational settings with the goal to enhance the learning process and increase student engagement, supported by the interactivity of new educational technologies [9]. *Gamification* is one of those methods.

2.2 Gamifying educational software

Gamification is generally defined as "the use of game design elements in non-game contexts" [10]. The goal of gamification in education is to enhance UX to boost student involvement with the use of appealing interface design elements and rewarding interactions with the software applications. Points, levels, progress bars, quests, badges, virtual goods, avatars, and leaderboards are some of the most popular gamification *mechanics* or *elements* [11]. In this study, points, badges, and leaderboards (also known as the PBL triad [12]) were selected due to their popularity in a range of domains, to be embedded in an educational application to evaluate their *individual* impact on UX and learning performance. Moreover, the three gamification elements were selected as they were the most suggested by (Ecuadorian and British) teachers and students answering to a pre-study interview, where participants were asked about their preferences regarding ludic and educational software.

Although the actual impact of gamification both in the short and long term is still debatable, research on the PBL elements has mostly found encouraging results. For example, research on leaderboards suggests that they can be a motivator of software use [13], whereas other studies have shown similar positive results for the use of badges [14] and points [15]. Therefore, in our research we study these three gamification elements separately, to assess their individual impact on UX and performance in students with different backgrounds. The decision to choose *gamification* as an engaging technique in this study is due to the challenging nature of the learning activity, as students were asked to synthetize their knowledge about a topic and create relevant questions (cf. Section 2.3) that were assessed to grant their quality with a reward.

2.3 Learning by questioning

Questioning can help develop critical thinking in many ways [16]. For example, allowing students to investigate a topic based on their own cross-examinations, enabling students to exchange and contrast ideas with other classmates, or allowing them to express their conceptual thoughts through verbal or written questions. In addition, the process empowers students to be more consciously involved in the learning process [17]. However, the act of creating quality questions is a cognitively demanding task as it requires in-depth information processing and a higher level of thinking [16, 18]. Hence, the use of *gamification* could encourage students through the task of questioning with the use of interactive and enjoyable interactions to sustain their mental effort for the task.

Interest in developing and studying questioning-based activities online appears to be growing, as technological tools are becoming more widely available for teachers and students to use. It is possible to see similar robust commercial software applications (e.g. Blackboard and Moodle) adapting their software for the integration of questioning activities and even digital rewards. Therefore, based on this growing interest, we investigate in this study the use of a gamified questioning-based online activity to support online learning in young students.

2.4 The digital divide

Worldwide, many teachers who have access to technology tend to incorporate it into their teaching practices, thereby enabling their students to have better access to learning resources. Among others, the benefits of educational technologies include promoting creativity, personal development, and satisfaction. However, large disparities are seen between the access and usage of educational technologies in countries from the Global North and the Global South [20] (cf. Section 2.1). Bridging this gap is therefore imperative to create fair information societies that can benefit from software applications regardless of their socio-economic disparities [21].

Moreover, understanding how socio-economic variables contribute to usage patterns could greatly improve the design of effective software systems that serve their original purpose, taking into consideration the particular needs of its users [22]. For example, as young people are now more connected than ever, this enables them to navigate at ease in most devices and platforms, whereas this is not the case in older user groups [23]. Digital skills and Internet use habits could be—in this sense—closely related to the economic, social, and cultural status of users. Disadvantaged students (such as many native from countries in the Global South, for instance) could potentially reach higher academic levels if they had better access to the Internet infrastructures and services. Consequently, governments, non-profit organisations, industries, and academia have done tremendous efforts in recent years to bridge the digital divide. Likewise, to provide fairer and more useful educational technologies, software designers have been turning to various inclusive practices such as conducting Participatory Design activities with key stakeholders from different socio-economic backgrounds (cf. Section 2.2), or designing software that could be easily adapted to different languages (cf. Section 4.5), benefiting students that otherwise would have not had access to such technologies.

3. METHODOLOGY

This study focuses on whether the use of specific gamification elements can have a significant impact on the user engagement, motivation, and performance of

lower secondary school students when learning about a physics topic using a questioning-based technique, and whether this impact is mediated by the participants' perceived usability of the software or their predisposition towards learning sciences. Our selection of this specific age group (12-16 years old) is due to the few studies that have been conducted with target groups other than those in Higher or Distance Education [11, 19]. The four versions of the educational software under evaluation are designated with acronyms for ease of reference: Non-Gamified - NG; Gamified with Points – GP; Gamified with Badges – GB; and Gamified with Leaderboards - GL.

3.1 Participants

This study was conducted with two secondary schools located in the same area of the highlands of Ecuador. **School Y** was a mixed gender secular school (with a neutral approach towards promoting religious content in their curricula). **School Z** was a single-sex school religiously affiliated (under the influence of the church, having religion as part of the school's curricula). Data of these two schools were analysed separately throughout this paper due to the significant differences found between the scores of School Y and School Z on the ATS, the SQM, and the initial knowledge test (pre-test) as shown in the sections below.

Schools Y and Z in Ecuador had access to fair scientific and technological facilities, compared to their counterparts Schools A and B in the United Kingdom which had very good technological facilities. School A was a public and secular secondary school, whereas School B was a private boarding school in the same geographical area [4].

3.2 Goals of the study

The core objectives of this study are:

(i) To contrast student perceptions of the gamification elements *points, badges, and leaderboards* implemented in Go-Lab.
(ii) To analyse if and how learning performance is affected with the integration of the gamification elements in the questioning-based activity prepared for this study.
(iii) To compare the user experience and the learning performance of secondary school students in two different countries.

3.3 Hypotheses

The *null* hypotheses of this study are closely related to the main goals presented in the section above.

H1a: There are no significant differences on the perceptions of software usability among students interacting with the NG, GP, GB, and GL versions of the software.
H1b: There are no significant differences on the self-reported perceptions of user engagement among the NG, GP, GB, and GL groups.
H1c: There are no significant differences on motivation among the NG, GP, GB, and GL groups.
H2a: There are no significant differences in the learning gains as operationalized by the difference between the pre- and the post-knowledge tests among groups.
H2b: There are no significant relationships between the students' pre-existing attitudes and motivations towards learning sciences, and their perception of software usability, their engagement with the activity, their motivation, and their learning gains.
H3a: There are no significant differences on the students' pre-existing attitudes and motivations towards learning sciences depending on the students' nationality.
H3b: There are no significant differences on the students' learning performance according to their country of origin.

3.4 Methods and Instruments

Several methods and instruments were used in this study to (i) test our null hypotheses and (ii) to collect data from the interactions of secondary school students with the educational software. These are summarised in the sections below.

3.4.1. Attitudes Towards Science Measures (ATS)
The ATS [24] measures the attitudinal predisposition of young students towards learning science. The ATS is composed of 34 items and six factors (Learning Science in School, Practical Work in Science, Science Outside of School, Importance of Science, Self-Concept in Science, and Future Participation in Science). Results are presented in this study on a range 0-1. The Cronbach's alpha is 0.872 in this sample, which indicates a high level of internal consistency (the original was a>0.700 [24]).

3.4.2. Science Motivation Questionnaire II (SMQ)
The SMQ [25] measures students' motivation to learn science in college and secondary school courses. It contains 25 items and four main factors (Intrinsic Motivation, Self-Determination, Self-Efficacy, and Career Motivation). Results are presented in this study on a range 0-1. The Cronbach's alpha in this sample is 0.776, which indicates a good level of internal consistency (the original ranged between 0.710 to 0.900 [25]).

3.4.3. Demographic Questionnaire
This is a homegrown questionnaire with questions relevant to the participants' background (gender, age), their self-reported measures of IT competence, their perceptions about the educational software they had used in the past, and their use habits and preferences.

3.4.4. Knowledge Tests (Pre- and Post-tests)
These are homegrown questionnaires comprising 12 questions. The pre- and the post-tests had identical questions, which were all multiple choice with a single possible correct answer.

3.4.5. Situational Motivation Scale (SIMS)

The SIMS [26] measures factors of situational intrinsic and extrinsic motivations, rated on a seven-point Likert scale. It comprises 16 items and four factors: Intrinsic Motivation, Identified Regulation, External Regulation, and Amotivation. The Cronbach's alpha in this sample is good, at 0.730 (the reported Cronbach's alpha was between 0.65 and 0.92 [26]).

3.4.6. System Usability Scale (SUS)
The SUS [27] measures subjective assessments of software usability. It is a ten-item questionnaire using a five-point Likert scale. Results are presented in a range 0-100. The Cronbach's alpha in this sample is good, at 0.760 (the reported Cronbach's alpha in past research was between 0.71 and 0.90).

3.4.7. User Engagement Scale (UES-SF)
The UES-SF questionnaire [28] measures subjective assessments of user engagement. It is a twelve-item questionnaire with four different factors (Focused Attention, Perceived Usability, Aesthetic Appeal and the Reward Factor). Results are presented on a scale 1-5. The Cronbach's alpha in this sample is 0.710 (the authors' was between 0.700 and 0.890 [28]).

3.4.8. Post-Intervention Survey
This survey was designed as an online poll (created with Kahoot, Quizizz, or similar) where students selected from multiple options what they considered more accurate to their case. Participants were asked for their feedback on their overall experience using the software application.

3.4.9. Group Discussion
Following the post-intervention survey, students were encouraged to exchange opinions through a group discussion moderated by a researcher. Group discussions were audio-recorded with the participants' consent during the experimental session of this study.

3.5 Other materials

Three other sources were used in this study: the online lesson that students received before completing any standardized questionnaires, the set of three questions students created about the topic of the online lesson, and the rewards students received for their effort in creating said questions.

3.5.1. The online lesson
Due to the schools' preferences and the students' age range, the subject of *electric circuits* was chosen by teachers from a predefined list of different scientific topics. An online lesson aligned with the relevant curricula of Ecuador and England was then selected from Go-Lab for the purpose of our studies. The online lesson (Figure 1) covered related concepts such as current, tension, and Ohm's law. In addition, the online lesson was adapted to enable *learning by questioning*. Phases of the activity were progressively made available to students according to the task they were asked to complete (e.g. read the online lesson, create questions, check rewards).

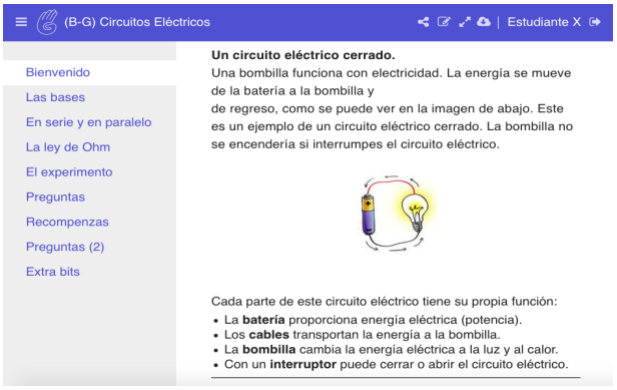


Figure 1: *Overview of the learning activity.*

3.5.2. Student questions
After finishing with the online lesson, students were asked to formulate three questions about the topic using the software. For this purpose, a new tab was made available for them on the environment of the activity (see *"Preguntas"* or *"Questions"* on the left-hand side menu on Figure 1). Student questions were assessed in this study manually by a researcher in terms of novelty, clarity, and relevance on a scale 0-5. *Novelty* referred to the level of interestingness and originality of the proposed question. *Clarity* referred to the understandability and clear and adequate structuring of the question. *Relevance* related to the closeness of the question to the topic of the learning activity.

3.5.3. The rewards
Three gamification elements—points (Figure 2), badges (Figure 3), and leaderboards (Figure 4)—were used to reward, in their respective group, the effort that individual students put in creating a set of questions about the topic of electric circuits. A non-gamified interface was shown to students in the control groups. Afterwards, students were asked to create three more questions for the chance to improve on their work (*"Preguntas 2"* on Figure 1).

Figure 2: *Example of the points reward.*

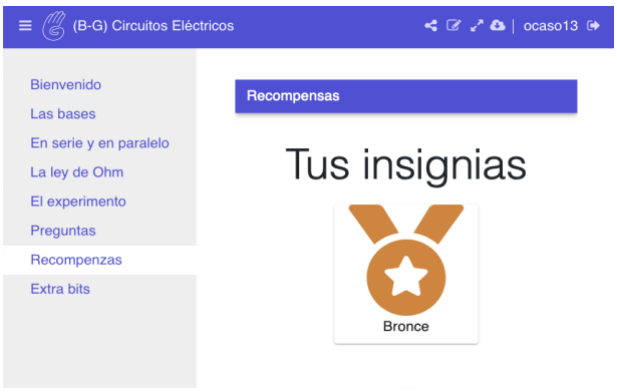


Figure 3: *Example of the badges reward.*

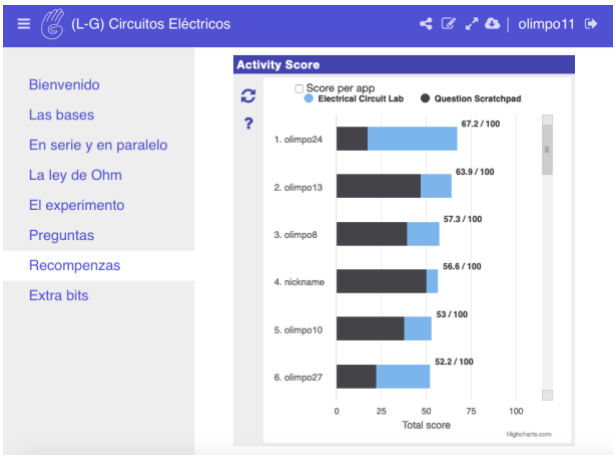


Figure 4: *Example of the leaderboard reward.*

3.6 Procedure

This study was designed as a questioning-based learning activity part of a physics lesson where the gamification elements were introduced individually within different groups to evaluate their impact on the student perception of software usability, their user experience (in terms of motivation and user engagement) and their learning performance. The study consisted of two experimental sessions with each participating class, in school premises where students had access to their own device. Prior the first session, participants gave appropriate consent and students filled in the ATS and SMQ under the supervision of their schools. Classes were randomly assigned to one of the four versions of the educational software (NG, GP, GB, GL).

During the first experimental session, students were asked to fill in a demographic questionnaire, they revised the same online physics lesson on the topic of Electric Circuits and completed a pre-test. Also, students were asked to create a set of three questions about the topic and they completed a motivational test. All students then had a break between 1-72 hours.

Starting the second session, students were presented with results to their three questions in the form of one of three gamification elements (points, badges, leaderboards). The non-gamified groups displayed a neutral interface of the learning activity, with no visible gamification element. Students were then given the opportunity to create a different set of three questions and they filled in a post-test. Afterwards, participants gave feedback about their perception of software usability (SUS), their user engagement (UES-SF), and for the second time, their motivation (SIMS). To wrap up the session, students answered an online survey and participated from a moderated group discussion.

4. RESULTS

Nine methods and instruments were applied to collect data from student participants in the age range of 12-16 years old from two schools in Ecuador. The main results from their responses and interactions are summarised in the sections below.

4.1 Demographic data

A total of 176 students participated in this study: 34 were male, 131 were female, and 11 people preferred not to say. The average student age in School Y was 15.31 (*SD=1.35*) and the average student age in School Z was 12.78 (*SD=0.63*). In School Y, 61.8% of students (N=42) claimed to spend 3-7 hours per day on an electronic device (0-3 hours= 17.6%; more than 7 hours= 20.6%). Similarly, the majority of participants in School Z (N=42, 38.9%) said to daily spend 3-7 hours on an electronic device (0-3 hours= 33.3%; more than 7 hours= 26.6%).

4.2 Pre-existent students' attitudes and motivations towards learning science

All 176 Ecuadorian students answered to the ATS (Section 3.4.1) and SMQ (Section 3.4.2) questionnaires prior the actual experimental sessions planned for this study (See Procedure, Section 3.6). On a scale 0-1, a total of 71 students from School Y reported their own attitudes towards science with a mean of 0.49 (*SD=0.27*), whereas 105 students in School Z gave it a mean of 0.68 (*SD=0.16*). On the same 0-1 scale, students in School Y reported their motivation to learn science with a mean of 0.53 (*SD=0.23*), while students in School Z had a mean of 0.71 (*SD=0.15*). Results of the Shapiro-Wilk test showed that data were not normally distributed, hence non-parametric tests were used. Results of the Mann-Whitney test showed significant differences between schools Y and Z for the ATS (U= 2298, Z= -4.404, p< 0.001) and the SMQ (U= 1949, Z= -5.538, p< 0.001).

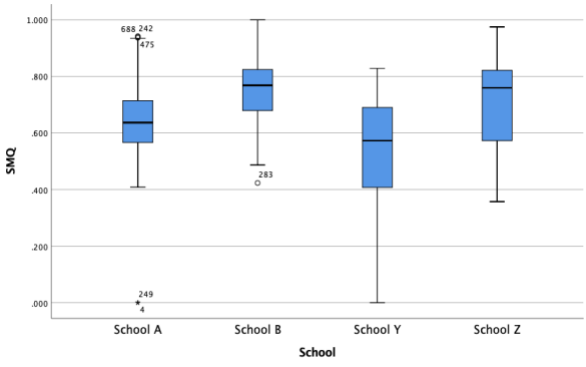


Figure 5: *SMQ scores in Ecuadorian and British Schools*

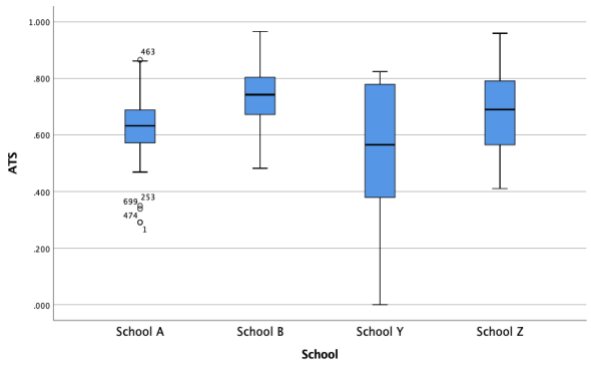


Figure 6: *ATS scores in Ecuadorian and British Schools*

In addition, results of the Mann-Whitney test showed significant differences between the Ecuadorian schools studied in this paper and British data previously collected [4], both for the ATS (*U*= 14579, *Z*= -2.884, *p*= 0.004) and the SMQ (*U*= 15310, *Z*= -2.044, *p*= 0.041), **rejecting** the **null hypothesis H3a**. Significant differences were found between the self-reported motivations (SMQ) and attitudes (ATS) of students prior the experimental sessions, depending on their country of origin (Figures 5 & 6).

4.3 Student situational motivation

Students were asked to complete the SIMS (Section 3.4.5) twice in this study, to measure any changes on their motivation during the learning activity. A total of 151 students filled in both questionnaires; 59 in School Y, and 92 in School Z (see Table 1 for details). Results of the Shapiro-Wilk test showed that data were normally distributed in all cases.

Overall, students had a slightly higher score on the second SIMS (*M*= 4.39, *SD*= 0.92) comparing it to the first SIMS (*M*= 4.38, *SD*= 0.80), but no significant differences were found between the two either in School Y (*t*(58)= -1.293, *p*= 0.201) or School Z (*t*(91)= 0.864, *p*= 0.390).

Table 1: *SIMS scores classified by school and group.*

SCHOOL/ GROUP		*N*	**SIMS1**		**SIMS2**	
			Mean	*SD*	*Mean*	*SD*
Y	*NG*	*17*	*4.43*	*0.89*	*4.56*	*0.80*
	GL	*15*	*4.36*	*0.74*	*4.33*	*1.06*
	GB	*12*	*4.58*	*0.74*	*4.85*	*1.05*
	GP	*15*	*4.48*	*0.69*	*4.66*	*0.82*
Z	*NG*	*26*	*4.56*	*0.95*	*4.56*	*0.94*
	GL	*26*	*4.36*	*0.84*	*4.42*	*0.97*
	GB	*20*	*4.13*	*0.91*	*3.72*	*0.77*
	GP	*20*	*4.19*	*0.49*	*4.19*	*0.60*

In School Y, results of the one-way ANOVA test showed insignificant differences on the motivational gains among groups (*F*(3,55)= 0.387, *p*= 0.762). In School Z, results of the ANOVA test also showed insignificant differences (*F*(3,88)= 1.530, *p*=0.212). These results **support** the **null hypothesis H1c.** No important changes were found on the motivation of the participants during the experiment.

Likewise, there were no statistically significant differences when analysing each of the four factors measured in the SIMS questionnaire (Intrinsic Motivation, Identified Regulation, External Regulation, Amotivation). In addition, no correlations were found between the pre-existent students' attitudes and motivations towards learning science and their situational motivation during the activity.

4.4 Software usability

Out of the 176 students participating in this study (see Table 2), 157 completed the SUS (Section 3.4.6). Overall, School Y (N=58) had a mean score of 61.29 (*SD=15.05*) and School Z (N=99) a mean score of 56.11 (*SD=14.50*). Data was normally distributed in School Y, but not in School Z.

Table 2: *SUS scores classified by school and group.*

Group	**School Y**			**School Z**		
	N	*Mean*	*SD*	*N*	*Mean*	*SD*
NG	*17*	*57.21*	*13.31*	*26*	*51.25*	*15.42*
GL	*15*	*64.50*	*11.06*	*26*	*57.50*	*13.93*
GB	*14*	*65.00*	*14.90*	*24*	*53.65*	*11.03*
GP	*12*	*58.75*	*20.82*	*23*	*62.61*	*15.44*

In School Y, results of the one-way ANOVA test showed insignificant differences among groups (*F*(3,54)= 1.044, *p*= 0.380). In School Z, results of the Kruskal-Wallis test showed significant differences among groups (*H*(3)= 9.928, *p*= 0.019). Results of Mann-Whitney tests in School Z showed a significant difference between the non-gamified group [NG] and the points group [GP] (*U*= 160.000, *Z*= -2.792, *p*= 0.005) with a medium effect size (Cohen's *d*=0.736, *r*=0.345). Based on these results, the **null hypothesis H1a** is **partially rejected.** Some differences were found on the perceived usability of the software application in School Z. No correlations were found between the perceptions of software usability and students' age, gender, or their previous attitudes and motivations to learn science.

4.5 User engagement

In total, 145 students (see Table 3) completed the UES-SF (Section 3.4.7). Overall, School Y (N=57) had a mean score of 3.43 (*SD=0.58*) and School Z (N=88) a mean score of 3.51 (*SD=0.68*). Results of Shapiro-Wilk tests showed that data were normally distributed in School Z, but not in School Y.

Table 3: *UES-SF scores classified by school and group.*

Group	**School Y**			**School Z**		
	N	*Mean*	*SD*	*N*	*Mean*	*SD*

NG	*17*	*3.30*	*0.46*	*17*	*3.62*	*0.68*
GL	*16*	*3.23*	*0.52*	*26*	*3.60*	*0.67*
GB	*13*	*3.89*	*0.52*	*22*	*3.31*	*0.63*
GP	*11*	*3.39*	*0.64*	*23*	*3.53*	*0.72*

Results of the Kruskal-Wallis test in School Y showed significant differences among groups ($H(3)= 9.928$, $p= 0.019$). Results of Mann-Whitney tests in School Y showed a significant difference between the non-gamified group [NG] and the badges group [GB] ($U= 43.000$, $Z= -2.830$, $p= 0.004$) with a large effect size (Cohen's $d=1.202$, $r=0.515$). In School Z, results of the one-way ANOVA test showed insignificant differences among groups ($F(3,84)= 0.917$, $p= 0.436$). Based on these results, the **null hypothesis H1b** is **partially rejected.** Some differences were found on the self-reported user engagement in the badges group (as compared to the control group) in School Y.

In addition to these results, linear regressions showed that the students' perceptions of software usability (SUS) had a significant influence on their user engagement (UES-SF) in both schools. More specifically, in School Y: ($F(1,54)= 16.683$, $p< 0.001$, $R^2= 2.4\%$); in School Z: ($F(1,86)= 34.856$, $p< 0.001$, $R^2= 2.8\%$).

4.6 Learning performance

As seen on Table 4, 163 students answered to both knowledge tests of this study (Section 3.4.4). Out of 12 possible points on the pre-test, School Y had a mean score of 5.74 (*SD=1.59*) and School Z a mean score of 4.86 (*SD=1.61*). On the post-test, School Y had a mean score of 6.26 (*SD=2.40*) and School Z a mean score of 5.69 (*SD=1.82*). Results of Shapiro-Wilk tests showed that data were normally distributed in all cases in School Z, but only in six out of eight cases in School Y.

Table 4: *Knowledge tests classified by school and group.*

SCHOOL/ GROUP		*N*	**Pre-test**		**Post-test**	
			Mean	*SD*	*Mean*	*SD*
Y	*NG*	*17*	*5.82*	*1.38*	*5.94*	*1.85*
	GL	*19*	*5.63*	*1.46*	*6.37*	*2.63*
	GB	*14*	*6.43*	*1.60*	*7.14*	*2.48*
	GP	*15*	*5.13*	*1.85*	*5.67*	*2.55*
Z	*NG*	*25*	*4.88*	*1.62*	*5.64*	*1.66*
	GL	*26*	*4.65*	*1.81*	*5.50*	*2.12*
	GB	*23*	*4.96*	*1.49*	*5.70*	*1.49*
	GP	*24*	*4.96*	*1.57*	*5.96*	*1.99*

In School Y, a paired sample t test showed a significant improvement on the learning performance of students ($t(64)= -2.030$, $p= 0.047$). However, when comparing the two knowledge tests using a Kruskal-Wallis test In School Y, insignificant differences were found among groups ($H(3)= 1.043$, $p= 0.791$). Likewise, in School Z a paired sample t test showed a significant improvement on the learning gains ($t(97)= -5.273$, $p< 0.001$), but results of the one-way ANOVA test showed insignificant differences among groups ($F(3,95)= 0.152$, $p=0.928$). These results **support** the **null hypothesis H2a.** No significant differences were found on the learning gains among groups in this study.

Additionally, linear regressions showed no influence of the students' previous attitudes (ATS) and motivations towards science (SMQ), their user engagement (UES-SF), and their perception of software usability (SUS) on the learning gains, **supporting** the **null hypothesis H2b**. Nonetheless, results of the one-way ANOVA ($F(1,340)= 4.839$, $p=0.028$) showed a significant difference with a low effect size (Cohen's $d= 0.241$, $r= 0.119$) on the learning gains of the Ecuadorian students studied in this report, compared to the learning gains of British students in a similar published study (see [4]) **rejecting** the **null hypothesis H3b.**

5. DISCUSSION

When comparing the outcomes of the Ecuadorian students reported in this study to the previously collected results of the British students (see [4]) significant differences were found between the self-reported *motivation* (SMQ, Figure 5) and *attitudes* of students (ATS, Figure 6) prior to the start of the experimental sessions (Section 4.2). For instance, as private schools with higher household incomes (School B in the UK and School Z in Ecuador) scored significantly higher in both scales than their counterparts in each respective country. This suggests that, depending on socio-economic factors, students could hold different views about science, which could consequently influence their predisposition to use educational software. However, as we are analysing merely two schools in each country, generalisations cannot be drawn from the results presented in this paper.

Concerning ***UX***, we see for example that, unlike the case of British secondary school students before studied [4], Ecuadorian students in this study did not show a pattern on their preferences towards the four versions of the software (see Table 2). The trend in British schools was to score the *usability* of the non-gamified version of the educational software the lowest, followed by the leaderboards, badges, and points, in that particular order (see Table 5). Nonetheless, we see that in both countries at least one school scored the usability of the software application gamified with *points* significantly higher than the non-gamified version, suggesting that the design of the gamification element, as well as how it was awarded and displayed was more usable to students. As points were awarded to all participants in the group (which was not the case with the badges, for example) another reason could be that students might have felt more attracted to the GP interface due to the novelty effect [29]. Overall, results from our studies suggested that students with a broader experience with educational software could have interacted with any version of the

software application at ease, without perceiving a significant difference in usability.

Additionally, we see that the *user engagement* was not affected similarly in both countries. In the United Kingdom (see [4]), students from both schools in the gamified groups were significantly more engaged than students in the non-gamified groups (Table 5). In this study, the Ecuadorian students interacting with the *badges* in School Y were the only group that showed a significantly higher user engagement than the control groups (Table 3). We believe this could be due to the previous exposure that participants from the British schools had similar reward schemes as part of their learning activities at school, which helped them recognise the attractiveness of receiving a digital reward. Also, we see that overall the Ecuadorian students rated their user engagement higher than British students, which could be due to—among other factors—the novelty effect [29], as students in the Global South commonly have less access to educational software to learn at school.

Table 5: *Usability (SUS) & User Engagement (UES-SF) scores in British schools classified by group.*

SCHOOL/ GROUP		**SUS**		**UES-SF**	
		N	*Mean*	*N*	*Mean*
A	*NG*	*25*	*50.00*	*25*	*2.52*
	GL	*23*	*50.75*	*20*	*2.66*
	GB	*19*	*52.37*	*19*	*2.98*
	GP	*26*	*63.46*	*26*	*3.07*
B	*NG*	*18*	*59.58*	*21*	*2.40*
	GL	*24*	*62.81*	*24*	*3.02*
	GB	*17*	*63.38*	*17*	*3.09*
	GP	*22*	*68.60*	*22*	*3.14*

Regarding students' *situational motivation*, both studies showed similar outcomes. An insignificant improvement was found in the motivation of Ecuadorian (Table 1) and British (Table 6) students which, we assume, could be due to the short interaction time students had with the software application, and also due to the topic of the online lesson which was mentioned several times by students during the group discussions in both countries and all schools. However, we see that the Ecuadorian scores on self-reported motivation are slightly higher than those of students in UK, which could be due to the novelty effect once again [29].

Table 6: *Situational Motivation (SIMS) scores in British schools classified by group.*

SCHOOL/ GROUP		*N*	**SIMS1**		**SIMS2**	
			Mean	*SD*	*Mean*	*SD*
A	*NG*	*24*	*3.73*	*0.61*	*3.71*	*0.60*
	GL	*21*	*3.83*	*1.05*	*3.78*	*1.05*
	GB	*26*	*3.30*	*1.02*	*3.57*	*0.91*
	GP	*28*	*3.85*	*0.61*	*3.88*	*0.93*
B	*NG*	*18*	*4.03*	*0.74*	*4.06*	*0.99*
	GL	*24*	*3.94*	*0.50*	*3.94*	*0.48*
	GB	*17*	*4.26*	*0.60*	*4.26*	*0.74*
	GP	*21*	*3.97*	*0.37*	*3.97*	*0.44*

Concerning ***learning performance***, The Ecuadorian students in this report showed insignificant differences in their learning gains among (NG, GL, GB, GP) groups (Table 4). On the other hand, results from the study with British students (see [4]) showed significant differences of the gamified groups compared to the non-gamified in both schools (Table 7). We suspected that one variable that could have played a pivotal role in the insignificant results of this study is the young age of students in School Z (see Section 4.1), who mentioned during the group discussions that they had very limited knowledge about the topic of *electric circuits.* Nonetheless, as the Ecuadorian students scored lower in the knowledge tests than the British students (suggesting a gap in their knowledge), they also had more room for improvement from the pre-test to the post-test, which could also help explain insignificant differences in this study.

Table 7: *Knowledge tests in British schools.*

SCHOOL/ GROUP		*N*	**Pre-test**		**Post-test**	
			Mean	*SD*	*Mean*	*SD*
A	*NG*	*24*	*6.33*	*0.48*	*5.88*	*0.41*
	GL	*22*	*6.50*	*0.55*	*7.73*	*0.40*
	GB	*23*	*6.26*	*0.44*	*6.83*	*0.55*
	GP	*26*	*8.88*	*0.42*	*8.96*	*0.43*
B	*NG*	*19*	*8.58*	*0.48*	*8.32*	*0.54*
	GL	*24*	*9.83*	*0.23*	*9.67*	*0.30*
	GB	*17*	*8.18*	*0.40*	*9.41*	*0.30*
	GP	*22*	*9.59*	*0.31*	*9.95*	*0.28*

Based on these empirical results, we infer that the design of the gamified questioning-based software could be improved for future studies, to benefit students in the Global North and the Global South more uniformly. Firstly, one should carefully select lesson topics closely aligned to global curricula, so that students on the same age range in any country could benefit from the use of the educational software equally. Secondly, by ensuring the consistency of the assessment criteria for determining the rewards to be granted, to avoid confusion in the participants. Thirdly, by using modern interactive technologies to aid the design of the gamification elements, that should appeal visually to the students with the use of appropriate colours, positioning, animations, etc. And, lastly, it could be beneficial to allow students to share questions and rewards with their peers to boost participation and engagement.

Nonetheless, answering to a post-intervention survey in this study, the vast majority of Ecuadorian students claimed to believe that technology could enhance their learning experience. More importantly, a total of 68.12% of students said that they would be more motivated to learn physics online if the software offered

a gamified style, compared to a 13.79% of students that said it would not make any difference to them (the rest were unsure or preferred not to answer).

In addition, students in both countries showed an increase in the quality of their questions, with a significant difference in the gamified groups as compared to the control groups (see Table 8). The mean of all questions created by Ecuadorian students during the first round increased from 3.11 (*N*= 169, *SD*= 0.50) to 3.54 points (*N*= 169, *SD*=0.59), whereas in the case of British schools it increased from 3.26 (*N*= 191, *SD*= 0.49) to 3.69 points (*N*= 191, *SD*= 0.54).

Table 8: *Quality of student questions classified by schools and groups.*

SCHOOL/ GROUP		*N*	**Q1**		**Q2**	
			Mean	*SD*	*Mean*	*SD*
A	*NG*	*25*	*3.13*	*0.42*	*3.31*	*0.54*
	GL	*21*	*3.14*	*0.56*	*3.56*	*0.51*
	GB	*29*	*3.16*	*0.51*	*3.72*	*0.56*
	GP	*27*	*3.16*	*0.53*	*3.56*	*0.54*
B	*NG*	*22*	*3.50*	*0.43*	*3.74*	*0.56*
	GL	*24*	*3.42*	*0.46*	*3.83*	*0.39*
	GB	*21*	*3.24*	*0.38*	*3.97*	*0.42*
	GP	*22*	*3.38*	*0.49*	*3.94*	*0.44*
Y	*NG*	*17*	*3.34*	*0.54*	*3.54*	*0.60*
	GL	*20*	*3.17*	*0.63*	*3.73*	*0.65*
	GB	*15*	*3.04*	*0.38*	*3.73*	*0.44*
	GP	*15*	*3.02*	*0.43*	*3.60*	*0.51*
Z	*NG*	*26*	*3.17*	*0.57*	*3.31*	*0.62*
	GL	*28*	*3.13*	*0.57*	*3.39*	*0.58*
	GB	*22*	*2.89*	*0.32*	*3.48*	*0.64*
	GP	*26*	*3.11*	*0.45*	*3.67*	*0.64*

However, due to the limited time of the experiment (around 90 minutes in total), it was not possible to measure the long-term changes on student behaviours and perceptions of the gamified questioning-based software. Future research could analyse the particular goals and needs of end-users in the Global North and the Global South to influence user enjoyment and long-term behavioural change, with the use of a wider range of gamification elements in a longitudinal study.

Likewise, focusing on a particular and relevant aspect of implementing gamification in educational software could enhance the interpretation of the results and strengthen the conclusions of future work. As this research involved several dimensions that were analysed individually and comparatively (software usability, user engagement and motivation, learning performance, quality of questioning, etc.), it was challenging to collect qualitative data on all aspects from the student participants in such a limited time. Therefore, future research should expand the exposure time students have with the gamification elements but focus on collecting enough data without harming students' willingness to participate, so they can consequently give proper feedback about their experience.

6. CONCLUSION

In this paper we contrast a sample of the Global North-South perspectives, to contribute to the understanding of how socio-economic factors could influence the effectiveness of online educational software applications. However, sociocultural differences are not systematically analysed in this study to avoid over-complicating the experiment and overburdening participants by collecting more data from them. As the small sample size is a limitation, no generalisation of the results of this work could be made, as it only compared two Ecuadorian schools with two British schools representing the Global South and Global North, respectively.

Nonetheless, based on the findings of our studies, we agree with some researchers [21, 22, 23] that have suggested that digital skills are related to the economic, social, and cultural status of students, as they depend on the access that they have to the Internet and educational online tools. Hence, governments should provide appropriate infrastructures, developers should consider inclusive design practices, and learning spaces should adequately fulfil their purposes, regardless of the environment in which they are being implemented.

Additionally, as also suggested by previous research [4, 11, 12, 13, 14, 15], results of this study showed that the individual implementation of the gamification elements points and badges had some positive effects on UX, particularly on the perceived usability of the software application and on user engagement (respectively, in this case). However, results of learning performance in students from the Global South did not correspond to the findings from the Global North, as all students in Ecuador benefited from the learning activity regardless of the group they were assigned to, whereas badges and leaderboards had significant effects on the learning gains in British students. Therefore, more qualitative research is needed to improve the understanding of how educational software impacts lower secondary school students from different socio-economic backgrounds, to support the development of effective educational software applications designed for this specific purpose and age group.

REFERENCES

[1] Carroll, J. M. (1997). Human-computer interaction: psychology as a science of design. Annual review of psychology, 48(1), 61-83.

[2] Charlotte Smith, R., Winschiers-Theophilus, H., Paula Kambunga, A., & Krishnamurthy, S. (2020, June). Decolonizing Participatory Design: Memory Making in Namibia. In Proceedings of the 16th

Participatory Design Conference 2020-Participation (s) Otherwise-Volume 1 (pp. 96-106).

[3] Wong-Villacres, M., Alvarado Garcia, A., & Tibau, J. (2020, April). Reflections from the classroom and beyond: Imagining a decolonized hci education. In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-14).

[4] Andrade, P., Law, E. L. C., Farah, J. C., & Gillet, D. (2020, December). Evaluating the effects of introducing three gamification elements in STEM educational software for secondary schools. In *32nd Australian Conference on Human-Computer Interaction* (pp. 220-232).

[5] Postman, N. (2011). Technopoly: The surrender of culture to technology. Vintage.

[6] Ons.gov.uk. 2020. Internet access – households and individuals, Great Britain: 2020. [online] Available at: <http://www.ons.gov.uk> [Accessed 10 April 2021].

[7] Ecuadorencifras.gob.ec. 2020. TIC 2018. [online] Available at: <www.ecuadorencifras.gob.ec/estadisticas/> [Accessed 10 April 2021].

[8] Choi, D., & Kim, J. (2004). Why people continue to play online games: In search of critical design factors to increase customer loyalty to online contents. CyberPsychology & behavior, 7(1), 11-24.

[9] Burbules, N. (2018). Watch IT: The risks and promises of information technologies for education. Routledge.

[10] Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification. using game-design elements in non-gaming contexts. In CHI'11 extended abstracts on human factors in computing systems (pp. 2425-2428).

[11] Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. Journal of Educational Technology & Society, 18(3), 75-88.

[12] Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. Computers in Human Behavior, 69, 371-380.

[13] Fotaris, P., Mastoras, T., Leinfellner, R., & Rosunally, Y. (2016). Climbing up the Leaderboard: An Empirical Study of Applying Gamification Techniques to a Computer Programming Class. Electronic Journal of e-learning, 14(2), 94-110.

[14] Denny, P. (2013, April). The effect of virtual achievements on student engagement. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 763-772).

[15] Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013). Disassembling gamification: the effects of points and meaning on user motivation and performance. In CHI'13 extended abstracts on human factors in computing systems (pp. 1137-1142).

[16] Bloom, B. S. (1956). Taxonomy of educational objectives. Vol. 1: Cognitive domain. New York: McKay, 20, 24.

[17] Campbell, J., & Mayer, R. E. (2009). Questioning as an instructional method: Does it affect learning from lectures?. Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 23(6), 747-759.

[18] King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. American Educational Research Journal, 27(4), 664-687.

[19] Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis.

[20] Generationunlimited.org. 2021. Digital Connectivity. [online] Available at: <https://www.generationunlimited.org/our-work/promising-ideas-innovation/digital-connectivity> [Accessed 16 April 2021].

[21] Martin, B. (2005). The information society and the digital divide: some North-South comparisons. International journal of education and development using ICT, 1(4), 30-41.

[22] Graafland, J. H. (2018). New technologies and 21st century children: Recent trends and outcomes.

[23] Gottschalk, F. (2019). Impacts of technology use on children: Exploring literature on the brain, cognition and well-being.

[24] Kind, P., Jones, K., & Barmby, P. (2007). Developing attitudes towards science measures. International journal of science education, 29(7), 871-893.

[25] Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. Journal of research in science teaching, 48(10), 1159-1176.

[26] Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). Motivation and emotion, 24(3), 175-213.

[27] Brooke, J. (1996). Sus: a “quick and dirty” usability. Usability evaluation in industry, 189.

[28] O’Brien, H. L., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. International Journal of Human-Computer Studies, 112, 28-39.

[29] Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). Does gamification work?--a literature review of empirical studies on gamification. In 2014 47th Hawaii international conference on system sciences (pp. 3025-3034). Ieee.

Towards Real-Life Adoption of Conversational Interfaces: Exploring the Challenges in Designing Chatbots That Live up to User Expectations

Luc van der Zandt
Drillster BV
De Meern, The Netherlands
luc@drillster.com

Esther van der Stappen
Avans University of Applied Sciences
Breda, The Netherlands
ej.vanderstappen@avans.nl

Koen van Turnhout
HU University of Applied Sciences
Utrecht, The Netherlands
koen.vanturnhout@hu.nl

Chatbots are increasingly popular, but state-of-the-art chatbots still struggle to meet user expectations, limiting their application in many domains. The factors affecting use have been studied extensively in laboratory contexts, resulting in context-independent requirements. However, user expectations and experiences of chat interfaces are affected by the context of use. Research efforts measuring experiences with chat interfaces need to shift from studies in controlled laboratory settings to studies in real-life settings in various domains. This paper explores this field of study by reporting on a small-scale real-life case study on the gap between expectations and experiences with an educational chatbot. More case studies in the wild, such as this one, could contribute to a deeper understanding of factors affecting acceptance and real use. We propose the use of the CIMO logic across these studies to build upon previous results.

Conversational User Interfaces, Educational Chatbots, Expectations–Experience Gap, Real-life, Case Study

1. INTRODUCTION

There is a long-standing tradition of research on natural language interaction with computers, dating back to the development of the famous ELIZA chatbot in the 1960s (Weizenhaum, 1966). This conversational interaction paradigm has had ups and downs over the years but is currently on the rise because of increased maturity of speech and language technology, the availability of speech-based assistants such as Alexa and Google Home on the market, and – particularly for text-based interfaces – the fact that the public is nowadays completely comfortable with communicating with short text messages such as with Facebook Messenger and WhatsApp (Dale, 2016). The design of conversational agents has been researched extensively. Recent studies focus, for example, on challenges like dialogue design and embodiment of chatbots (Foster, 2019; Fischer et al., 2019), specific design requirements for different roles and domains of chatbots such as comedy (Perone and Edwards, 2019), news delivery (Dubiel, Cervone and Riccardi, 2019), therapy and tutoring (George, 2019), and safeguarding qualities in interaction with chatbots such as trust (Rheu et al., 2020) and engagement (Candello et al., 2019).

However, as conversational technology is maturing and applied more and more in real-life contexts, it has become evident that a mismatch between user expectations and experience exists, limiting user adoption. This notion became mainstream in the field of study to conversational agents since the 2016 publication by Luger and Sellen (2016), which mentions a 'gulf of evaluation' for conversational agents. Luger and Sellen found that users constructed poor mental models about the workings of the interface and, as such, set unrealistic expectations for it. This mismatch between the user expectations and the actual system may be particularly urgent in language interfaces. Unlike graphical user interfaces, they do not convey much information about their action possibilities in their appearance.

Later studies in controlled laboratory settings have focused on discovering more concrete factors affecting use and generic requirements for developing chat interfaces. Weber and Ludwig (2020), for example, identified user needs for chatbots. Among others, they elaborate on the need for initial explanation and guidance, natural conversation, the ability of the agent to remember the context of the conversation, customisation of the interaction, alternative ways of contact, technical robustness, and keeping in control of if, and when, the agent proactively contacts the user through (push) notifications. In turn, Følstad et al. (2018a) focused on the factors specifically influencing trust in conversational agents for customer service. This study found that users value a chatbot's ability to understand the user, human-likeness, self-preservation, professional appearance, and security and privacy. As the authors of these studies acknowledge, such lab studies are suitable for uncovering generic requirements for conversational interfaces, but they lack sensitivity for needs that arise out of the context of use. This usage context is expected to be an important determinant of user expectation which warrants case studies that particularly explore those contexts. Mapping expectations and experiences of chat interfaces in those contexts may bring forward different requirements and priorities than those previously conceived. As such, this paper argues that research efforts measuring experiences with chat interfaces need to shift from studies in controlled laboratory settings to studies in real-life settings from a variety of domains.

The application of chat interfaces in real-life settings in the educational domain is a particularly unexplored area in the existing literature which may hold perspectives for future design efforts within and outside this domain. This paper explores this field of study by reporting on a small-scale real-life case study on the gap between expectations and experiences with an educational chatbot. In the following sections, we describe the used methods and our findings. We end the paper with a call for further real-life case studies and propose a method for knowledge-building across these cases.

2. METHODS

In the case study described in this paper, we realised and evaluated a chat interface to support learning using an existing adaptive learning system called Drillster. Drillster is a question-based adaptive learning tool. It uses a proprietary algorithm that incorporates elements of graduated interval recall (Pimsleur, 1967) and the Leitner system (Leitner, 1972) to battle the 'forgetting curve', a term coined by Hermann Ebbinghaus (1885) to describe the gradual decline of memory retention over time. Users can create, share and do exercises called 'drills' to gain and retain knowledge. The company states that "incorrectly answered questions are repeated more often than correctly answered ones" (Drillster BV, 2019). The platform predicts the forgetting curve of the user and, based on this data, brushes up the knowledge that is likely to be forgotten soon. A chat interface would support users by actively reminding and involving them to gain and retain knowledge via a familiar, highly used interface. After its realisation, we introduced the interface in a real-life educational setting to 33 Dutch high school students aged 14 to 18, who used and evaluated it in their daily educational tasks in the school subjects Greek, Latin and biology over two weeks. Prior to the intervention, the participants already used Drillster for their education via the conventional interfaces, i.e., web and mobile.

2.1 Gathering User Expectations

To determine what these potential users expected from a chat interface for the Drillster platform, we provided our subjects with a survey of four open questions asking how the subjects thought about a chatbot for the Drillster platform, what positives and negatives they expected from it, and what effect they expected that the interface would have on their frequency of use. This approach provided us with insight into the expectations and wishes of the subjects but did not establish specific functional requirements, as the functional outline was determined by the learning platform and were beyond the scope of this study. We asked the following four questions (translated from Dutch):

- What are your thoughts on a chatbot to practise drills? Please write down what first comes to mind.
- What are your hopes on what a chatbot can do for you?
- What are your hopes on what a chatbot will certainly not do?
- Do you think you will use Drillster more active, less active, or with an equal frequency because of a chatbot? Why?

Using deductive thematic analysis as described by Braun and Clarke (2012), we coded the responses to these questions and categorised them into themes. The approach describes six phases: familiarisation with the data, generating initial codes, searching for themes, reviewing potential themes, defining and naming themes, and producing the report.

2.2 Gathering User Experiences

After initial familiarisation with the users and their expectations, a large part of the study consisted of designing and realising the prototype. The prototype evolved in iterations with the study's knowledge acquisition, using an action design-based approach (Sein et al., 2011). When the prototype reached a usable state, we asked the participants to use the prototype for two weeks. This comprised practising drills for the concerning subject that were pre-made by the teacher. Consequently, we evaluated it using a questionnaire constructed with the modular extension of the User Experience Questionnaire (UEQ+) (Laugwitz et al., 2008, Schrepp and Thomaschewski, 2019). We could map the user expectations discovered by our thematic analysis to the actual experience found by analysing responses to the UEQ+. The evaluation used the following nine scales: attractiveness, efficiency, perspicuity, dependability, stimulation, novelty, trust, usefulness, and intuitive use, which could be evaluated on a scale from -3 to +3. We chose the UEQ+ in favour of other models such as TAM (Davis, 1989), TAM2 Venkatesh and Davis, 2000), and UTAUT (Venkatesh et al., 2003), because these instruments include external (organisational) factors influencing actual usage. The present study mainly focuses on intrinsic user intentions and motivations, making a questionnaire solely focusing on these factors more appropriate. Therefore, we advocate using the User Experience Questionnaire (UEQ) and its modular extension, the UEQ+, in the present study.

3. RESULTS

To evaluate our prototype, we asked students to use the prototype for two weeks to learn for school subjects that already incorporated Drillster – Greek, Latin and Biology. After these two weeks, we asked the students to evaluate the prototype using the UEQ+. Twenty-four responses came in (N=24). We detected major inconsistencies among the answers to the items in the scale 'trust' (α=.1). Because of this, we omitted the scale' trust' from the results. Careful comparison between the pre-test and post-test results provided insight into the prototype's attributes on which the expectations and experience differed the most.

3.1 An Efficient, Fast and Stimulating Interface

The results indicate a distinct difference between expectations and experience on efficiency and stimulation. In the general question about expectations in the pre-test, respondents named these factors as a positive, more so than the other themes. Furthermore, a total of 22 out of 39 discovered codes on the question *What are your hopes on what a chatbot for the Drillster platform can do for you?* were categorised as positive expectations regarding 'efficiency' and 'stimulation'. However, the efficiency and stimulation scales received the two lowest mean evaluations in the post-test ($\bar{x}$=.48 and $\bar{x}$=.47). Participants particularly leaned toward rating the application as slow ($\bar{x}$=.25), while the pre-test showed they hoped the chatbot would help them learn more efficiently (6/16 codes categorised in 'stimulation') and feared it would be slow (6/12 codes categorised in 'dependability'). These results indicate that students had high expectations regarding the chatbot's effects on their learning efficiency and stimulation, but the chatbot could not live up to them.

3.2 Trust and Transparency

Looking at the pre-test results, a big concern of respondents was that the chatbot would send spam messages or unsolicited notifications (15/36 codes on the question *What are your hopes on what a chatbot will certainly not do?*). Furthermore, some feared it would invade their privacy (2/36 codes) and be slow or dumb (12/36 codes). We could not accurately measure the 'trust' scale in the post-test, but the pre-test results indicate that our implementation should comfort the user regarding spam messages and data access and answer the user's queries as fast as possible.

3.3 Straightforward Navigation and Overview of Capabilities

Furthermore, the pre-test responses indicated that respondents hoped the chatbot would enable easy and accessible learning (10/39 codes on the question *What are your hopes on what a chatbot can do for you?*). While the chatbot's perspicuity scored relatively well in the post-test ($\bar{x}$=1.15), the fact that it was such a prominent positive expectation for the chatbot in the pre-test indicates that perspicuity is an important factor influencing the user experience of our prototype. Improvements regarding perspicuity may be beneficial to the user experience and acceptance.

3.4 Interface Flexibility

Apart from the fear that a chatbot would be slow, respondents also indicated that they feared a chatbot would need specific input. We can categorise responses into two categories: (1) needing specific input for chatbot instructions and (2) needing specific answers to practice questions. The need for user error correction seems particularly apparent for text-based interfaces, such as chatbots, where typos can easily be made. While respondents positively evaluated the chatbot's dependability ($\bar{x}$=1.35) in the post-test evaluation, anticipating user errors by correcting their input to the extent that the learning effect is not affected may improve the acceptance of our prototype.

3.5 An Authentic Messaging Experience

Lastly, our results indicate that students open and use messaging apps more often than a Drillster client application and, thus, would be motivated to use a Drillster interface on such apps more often than the usual Drillster client applications (18/33 responses on the question *Do you think you will use Drillster more active, less active, or with an equal frequency because of a chatbot?*). However, the post-test evaluation results indicate that the use of the chatbot did not feel intuitive, as the mean evaluation of the scale 'intuitive use' is the fourth-lowest rated scale in the post-test ($\bar{x}$=.9). This suggests the interface did not quite feel like an authentic messaging experience, and the integration with the messaging platform could have been more seamless.

4. CONCLUSIONS AND DISCUSSION

The present study reports on a small-scale real-life case study of an educational chatbot. It aims to contribute to a deeper understanding of the expectations and experiences of chatbot users in real-life settings. We elaborate on the scarce existing work regarding motivations to use chatbots in the wild and draw conclusions regarding the gap between expectations and experience observed in this case study in relation to controlled laboratory studies and previous studies regarding general-use and customer service chatbots. For this, we realised a chatbot for an e-learning platform and asked a group of 33 high school students about their expectations and asked them to evaluate the application after using it, for which 24 responses came in.

4.1 Discussion of Results

Our findings align well with existing work. We found that respondents did not really know what to expect but generally had an overall positive attitude towards the idea of a chatbot. Next to stimulation to learn, respondents indicated they hoped the chatbot would benefit their efficiency by being able to learn easier and faster. This sentiment of a chatbot aiding general task efficiency is shared by Følstad et al. (2018b, p. 7) and Lugger and Sellen (2016). However, these studies did not find motivations regarding stimulation of use, which is a big motivator in the present study. We could explain this difference because our case is in the educational domain, where task motivation can generally not be taken for granted. The cases of Følstad et al. and Lugger and Sellen focus on chat assistants in the general domain (i.e., to manage everyday tasks) and customer service, respectively. In these cases, task motivation may be more present or even be taken for granted.

We found that respondents feared the chatbot would send spam messages, a finding also reported by Weber and Ludwig (2020, p. 325). Respondents indicated they feared the chatbot would not understand their input, which was also called a challenge by Følstad et al. (2018a). This study also named straightforwardness as a perceived benefit, but we did not discover this in the present study. We could explain this discrepancy by the difference between chatbots' application in customer service, on which Følstad et al. focused, and the educational domain.

While the responses to the questions on expectations indicated students expected the chatbot would benefit their efficiency and stimulation, these scales were rated lowest in the subsequent evaluation. This finding suggests that a substantial gap exists between expectations for, and experience with, conversational user interfaces within the real-life context of our case study. This finding matches that of Luger and Sellen (2016), who found users failed to construct adequate mental models of the intelligence, capabilities, and goals of conversational agents in the general domain. In our case study, this seems to be focused on the capability of the chatbot to improve learning efficiency and stimulation, as a motivator to use a chatbot is to study faster, i.e., more efficiently.

4.2 Directions for Future Research

This study started to explore the design space of chatbots in real-life contexts by realising and evaluating an educational chatbot. While we believe our findings can, to an extent, be valuable to future design efforts, the validity of the results of this single case study with few respondents is limited. More case studies in the wild, such as this one, could contribute

to a better understanding of context-dependent factors affecting the real use of chat interfaces.

For effective knowledge build-up across case studies, it is important to be very explicit about the scope of individual findings. Therefore, we propose using the CIMO logic (Denyer et al., 2008) to construct design propositions for future design and research efforts. The logic is as follows: "in this class of problematic contexts (C), use this intervention (I) type to invoke these generative mechanism(s) (M), to deliver these outcome(s) (O)" (Denyer et al., 2008).

We argue that using the CIMO logic in HCI holds several benefits. The components of the logic form a specification that can help others identify the applicability of the proposition to their case, making it easier to attribute different findings in different studies to a specific element of the CIMO specification. This is particularly valuable for knowledge build-up across multiple field studies where the precision of CIMO aids in diagnosing conflicting findings, leading to a refinement of the specificity, practical applicability, and robustness of the propositions; elements which are needed for all forms of prescriptive knowledge (van Turnhout et al., 2019)**.**

New research and development challenges anchored in the contexts of use arise when novel technologies such as conversational interfaces reach maturity, sparking adoption in the real world. It is of key importance that we explore those situated challenges using field studies, like the one reported in this paper.

5. REFERENCES

Braun, V. and Clarke, V. (2012) Thematic Analysis. In: Cooper, H. et al. (eds). *APA handbooks in psychology. APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological*. Washington: American Psychological Association.

Candello, H. et al. (2019) Can Direct Address Affect Engagement with Chatbots Embodied in Physical Spaces? In: Cowan, B.R. and Clark, L. (eds). *CUI '19: Proceedings of the 1st International Conference on Conversational User Interfaces, Dublin, August 2019.* New York: ACM. 1–3.

Davis, F.D. (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13 (3), 319–340.

Denyer, D., Tranfield, D. and Aken, J.E. van (2008) Developing Design Propositions through Research Synthesis. *Organisational studies*, 29 (3). 393–413.

Drillster BV (2019) *The Science Behind Drillster*. Available from https://drillster.com/wp-content/uploads/2019/08/White-Paper-_The-science-behind-Drillster_.pdf (24 October 2020).

Dubiel, M., Ceryone, A. and Riccardi, G. (2019) Inquisitive mind: a conversational news companion. In: Cowan, B.R. and Clark, L. (eds). *CUI '19: Proceedings of the 1st International Conference on Conversational User Interfaces, Dublin, August 2019.* New York: ACM. 1–3.

Ebbinghaus, H. (1885) Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie. Leipzich: Decker & Humblot.

Fischer, J.E. et al. (2019) Progressivity for voice interface design. *Proceedings of the 1st International Conference on Conversational User Interfaces*. *Dublin, August 2019.* New York: ACM. 1–8.

Følstad, A., Nordheim, C.B. and Björkli, C.A. (2018a) What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In: Bodrunova, S. et al. (eds). *Internet Science. INSCI 2018. Lecture Notes in Computer Science, vol 11193*. Cham: Springer.

Følstad, A., Skjuve, M. and Brandtzaeg, P.B. (2018b) Different Chatbots for Different Purposes: Towards a Typology of Chatbots to Understand Interaction Design. In: Bodrunova, S. et al. (eds). *Internet Science. INSCI 2018. Lecture Notes in Computer Science, vol 11551*. Cham: Springer.

Foster, M.E. (2019) Face-to-Face Conversation: Why Embodiment Matters for Conversational User Interfaces. In: Cowan, B.R. and Clark, L. (eds). *CUI '19: Proceedings of the 1st International Conference on Conversational User Interfaces, Dublin, August 2019.* New York: ACM. 1–3.

George, S. (2019) From sex and therapy bots to virtual assistants and tutors: how emotional should artificially intelligent agents be? In: Cowan, B.R. and Clark, L. (eds). *CUI '19: Proceedings of the 1st International Conference on Conversational User Interfaces, Dublin, August 2019.* New York: ACM. 1–3.

Laugwitz, B., Held, T. and Schrepp, M. (2008) Construction and Evaluation of a User Experience Questionnaire. In: Holzinger, A. (eds). *HCI and Usability for Education and Work. USAB 2008. Lecture Notes in Computer Science, vol 5298.* Heidelberg: Springer. 63–76.

Leitner, S. (1972) *So lernt man lernen. Der Weg zum Erfolg.* Freiburg im Breisgau: Herder.

Luger, E. and Sellen, A. (2016) "Like Having a Really Bad PA": The Gulf Between User Expectation and Experience of Conversational Agents. In: Kaye, J. and Druin, A. (eds). *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. San Jose, May 2016.* New York: ACM. 5286–5297.

Pimsleur, P. (1967) A Memory Schedule. *Modern Language Journal*, 51 (2). 73–75.

Rheu, M. et al. (2020) Systematic Review: Trust-Building Factors and Implications for Conversational Agent Design. *International Journal of Human-Computer Interaction*, 37 (4). 1–16.

Roda, C., Angehrn, A.A. and Nabeth, T. (2001) Conversational Agents for Advanced Learning: Applications and Research. *Proceedings of BotShow 2001 Conference*. *Fontainebleau, 2001.* Fontainebleau: INSEAD. 1–7.

Schrepp, M. (2018) *UEQ_Plus_Data_Analysis_Tool.xlsx*. Available from http://ueqplus.ueq-research.org/Material/UEQ_Plus_Data_Analysis_Tool.xlsx (18 December 2020).

Schrepp, M. and Thomaschewski, T. (2019) *Handbook for the modular extension of the User Experience Questionnaire*. Available from http://ueqplus.ueq-research.org/Material/UEQ+_Handbook_V1.pdf (18 December 2020).

Sein, M.K. et al. (2011) Action Design Research. *MIS Quarterly*, 35 (1). 37–56.

Turnhout, K. van et al. (2019) A Practical Take on Theory in HCI. White paper.

Venkatesh, V. and Davis, F.D. (2000) A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, 46 (2). 186–204.

Venkatesh, V. et al. (2003) User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27 (3). 425–478.

Weber, P. and Ludwig, T. (2020) (Non-)Interacting with conversational agents: perceptions and motivations of using chatbots and voice assistants. *Proceedings of the Conference on Mensch und Computer*, *Magdeburg, September 2020.* New York: ACM. 321–331.

Weizenbaum, J. (1966) ELIZA–a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9 (1). 36–45.

Paper session 8: Socio-Technical HCI. Session chair: Christina Li

Vibe or Light? Someone or All?: Effects of Feedback Modality and Who Receives Feedback on Meeting Support

Junko Ichino
Tokyo City University
Yokohama, JAPAN
ichino@tcu.ac.jp

Yoshiko Yagi
ITOKI CORPORATION
Tokyo, JAPAN
yagi9yw5@itoki.jp

Teru Ozawa
ITOKI CORPORATION
Tokyo, Japan
ozawa8q4a@itoki.jp

We explore how meeting members modify their responses to feedback according to the feedback modality and who receives the feedback. We conducted a field study and compared four feedback conditions: three using vibrotactile modality (chair vibration) and one using visual modality (spotlight flashing). The three vibrotactile conditions differ in the feedback recipients: potential speaker (a member whom other members would like to hear speak next, or a member who is willing to speak next), current speaker, and all members. Regarding the modality, the vibrotactile modality provided a moderate level of distraction of members (while the visual modality was low enough to be ignored) and led to more turn-taking than the visual modality. Regarding the recipients, members felt more positively about feedback when potential speaker, rather than current speaker, received feedback. Also members resulted in more turn-taking when all members or current speaker, rather than potential speaker, received feedback.

Vibrotactile. Feedback modality. Feedback recipient. Meeting support. Discussion support. Conversation support.

1. INTRODUCTION

People relate differently to computers with different interfaces (Reeves and Nass (1996)). In future, when people and highly intelligent computers coexist symbiotically, how does the interface of such highly intelligent computers impact how people relate to computers and how people conduct their intellectual and creative activities? Meetings is an example of an intellectual and creative activity of people. Fifty percent of meeting time is unproductive and twenty-five percent is spent on topics unrelated to the agenda (Doyle and Straus (1993)). So that means it is not straightforward for meeting members to effectively conduct meetings by themselves (Doyle and Straus (1993)). Therefore, there has long been a strong demand for effective support for facilitating meetings using computers.

Previous work on meeting support using computers has explored how the computer interface impacts meeting members' responses to computers and how it impacts their meetings. In most existing systems, computers provide feedback on how actively each member has so far participated in a meeting (i) using the visual modality (e.g., DiMicco et al. (2004); Bergstrom et al. (2007); Kim et al. (2008); Nowak et al. (2012)), and (ii) to all members (e.g., (DiMicco et al. (2004); Bergstrom et al. (2007); Kim et al. (2008); Balaam et al. (2011)), to encourage or discourage their participation in the meeting. However, this approach does not necessarily lead to satisfactory results: (i) visual feedback sometimes distracts members, and (ii) providing feedback to all members sometimes makes members who are less active in their participation feel negatively about feedback.

In this paper, we consider meeting support using computers, and explore how meeting members modify their responses to feedback depending on two key factors of the interfaces: the modality used for feedback ((i) above) and which members receive feedback ((ii) above). To fully explore this, computers for meeting support must be highly intelligent and comparable to a human facilitator; however, such computers are beyond what is currently available. In addition, human social behavior is difficult to reproduce in a laboratory experiment. We therefore employ a field study using the Wizard of Oz method with a human facilitator. We obtained the cooperation of 17 office workers in a company to conduct a field study of their actual brainstorming sessions. In our field study, we considered two modalities used for feedback and three subsets of members to receive the feedback. Specifically, we compared four feedback conditions: three using the vibrotactile modality and one using the visual modality. The feedback conditions using the vibrotactile modality vibrate a chair of each member, and the feedback condition using the visual modality flashes a spotlight on the ceiling. The three vibrotactile feedback conditions differ in which members receive feedback: (1) a member whom other members would like to hear speak next, or a member who is willing to speak next (a potential speaker), (2) a member who is currently speaking (a current speaker), and (3) all members.

2. RELATED WORK

There exists much work that addresses supporting meetings using computers and explores how such

support impacts the responses of meeting members to computers. Many existing meeting support systems monitor either verbal (e.g., Leshed et al. (2009); Tausczik and Pennebaker (2013)) or nonverbal (e.g., Balaam et al. (2011); Nowak et al. (2012); Sanchez-Cortes et al. (2012)) communication among members and provide real-time feedback to members regarding specific aspects of their communication (e.g., Soller et al. (2005)). In this section, we analyze previous work according to three criteria.

The first criterion is feedback modality ((i) in section 1). Many meeting support systems use the visual modality (such as graphics (e.g., Bergstrom et al. (2007); Kim et al. (2008); Terken and Sturm (2010); Nowak et al. (2012)), animation (e.g., Leshed et al. (2009); Balaam et al. (2011); Schiavo et al. (2014); Tausch et al. (2016)), text (e.g., Leshed et al. (2009); Schiavo et al. (2014)), and light (e.g., Bachour et al. (2010); Occhialini et al. (2011); Snyder et al. (2015); Dagan et al. (2019)) for feedback to meeting members. However, the visual modality sometimes distract members from their primary activities. To avoid this shortcoming, some studies have explored ambient or peripheral interfaces (e.g., Streng et al. (2009); Occhialini et al. (2011); Tausch et al. (2014); Snyder et al. (2015)). However, even with these interfaces, some members still feel distracted by visual feedback (e.g., Leshed et al. (2009); Balaam et al. (2011); Schiavo et al. (2014)). In contrast, the vibrotactile modality is promising: it is available even when a user's auditory and visual modalities are occupied or limited, because of the user's primary activities or because of social and environmental factors, and it is expected to be less distracting to users. For example, Pielot et al. (Pielot et al. (2013)) showed that feedback using the vibrotactile modality does not distract users when the "right" vibration patterns are used. Many existing systems exploit the benefits of the vibrotactile modality by using tactile stimuli, such as vibration, and providing information on wearable devices (Birnholtz et al. (2015); Schneider et al. (2016)), such as smart phones (Pielot et al. (2013); Saket et al. (2013)), smart watches (Blum and Cooperstock (2016)), and shoes (Meier et al. (2015)). Therefore, we believe that vibrotactile feedback would not significantly distract members and would effectively support communication activities, such as meetings, that fully occupy members' visual and auditory senses. Few previous work has explored this possibility. Dagan et al. (Dagan et al. (2018); Dagan et al. (2019)) proposed an interface that provided haptic (i.e., vibrotactile) feedback in group discussions and reported their preliminary findings.

The second criterion is who receives feedback ((ii) in section 1). Many meeting support systems provide feedback to all members (e.g., DiMicco et al. (2004); Bergstrom et al. (2007); Kim et al. (2008); Bachour et al. (2010)). Members, who are less active in their participation, using such systems are reported to sometimes feel negatively about feedback; for instance, they feel frustrated about other members also receiving feedback and, knowing how actively they have participated in the meeting, feel forced to participate, and they feel alienated from being unable to participate fully (Bachour et al. (2010); Schiavo et al. (2014); Tausch et al. (2016)). This suggests that providing feedback to all members may not be appropriate for less active members. When some, not all, members receive feedback, how would members, especially those who are less active, feel about feedback? Previous work has not explored this.

The third criterion is whether feedback effectively facilitates meetings. How do members modify their responses to feedback when the feedback effectively facilitates meetings? The basic principle of previous work is to provide feedback on group dynamics (the balance of participation, primarily in speaking activities, among members; for example, how often and how long each member has so far spoken in the meeting (e.g., DiMicco et al. (2004); Bergstrom et al. (2007); Kim et al. (2008); Streng et al. (2009); Bachour et al. (2010); Tausch et al. (2016)), to make members aware of their group dynamics, and consequently to lead members to voluntarily modify their behavior (Tausczik et al. (2013)). Previous work has adopted objective measures, such as the probability of turn-taking (e.g., Kim et al. (2008); Terken and Sturm (2010)), balance of member participation (e.g., DiMicco et al. (2004); Kim et al. (2008); Bachour et al. (2010); Tausch et al. (2016)) and of types of remarks (e.g., Leshed et al. (2009); Snyder et al. (2015)), and members' subjective evaluation of these measures (e.g., DiMicco et al. (2004); Bachour et al. (2010); Nowak et al. (2012); Tausch et al. (2016)). In some systems, individual members voluntarily modified their behavior in response to feedback, resulting in improved group dynamics (e.g., DiMicco et al. (2004); Tausczik and Pennebaker (2013); Tausch et al. (2016)). However, in some systems, less active members did not necessarily increase their participation, although more active members often decreased their participation (e.g., DiMicco et al. (2004); Bachour et al. (2010)). No previous work has explored whether feedback effectively facilitates meetings, considering feedback modality and who receives feedback, as described above.

3. RESEARCH QUESTIONS

We pose the research questions based on the above three criteria. Feedback modality (criterion 1) relates to RQ1, and feedback recipients (criterion 2) relates to RQ2. Facilitation of meetings (criterion 3) considering the modality relates to RQ1–2, and considering the recipients relates to RQ2–2.

RQ1: How do meeting members modify their responses to feedback when different modalities (vibrotactile and visual) are used for feedback? (Feedback modality)
RQ1–1: How are meeting members distracted when different modalities are used for feedback?
RQ1–2: How are meetings facilitated when different modalities are used for feedback?
RQ2: How do meeting members, especially those who are less active, modify their responses to feedback when different subsets of members receive feedback? (Feedback Recipients)
RQ2–1: How do meeting members feel about feedback when different subsets of members receive feedback?
RQ2–2: How are meetings facilitated when different subsets of members receive feedback?

4. FEEDBACK DESIGN

To simplify our analysis of how members modify their responses to feedback, we only use feedback in its simplest form. Specifically, feedback carries a simple message: either "please speak" or "please encourage someone else to speak."

4.1 Comparison with vibrotactile feedback

We compare the visual modality with the vibrotactile modality, as a number of previous work used the visual modality (i.e., "baseline"). We adopt light (room illumination) to implement feedback using the visual modality. Using light as feedback allows members to focus on their primary activities without being distracted, and so it is fair to use it in comparison with vibrotactile feedback.

4.2 Feedback Pattern

We design a feedback pattern that less distract meeting members.

Regarding vibration, we apply the guideline for designing vibration-based interfaces (Saket et al. (2013)). They found that three factors contributed to a user's perceived urgency of vibration alerts: the gap length between vibrations, the number of gaps, and the vibration length. They also found that the pattern of "short on (vibration) and long off (gap)" was perceived as the least urgent. We argue that a lower level of perceived urgency causes less distraction to the members, and we apply their findings to our study. In our a preliminary study, we varied the on and off lengths of "short on and long off" in a vibration pattern and chose the pattern of "one second vibration on and two seconds vibration off" to avoid distraction of members. To ensure that the members notice a feedback, we decided to use the pattern of three consecutive pairs of "one second on and two seconds off."

Regarding light, we conducted a preliminary study and decided to use the pattern of three consecutive pairs of "one second light on and two seconds light off," which is similar to the vibration.

4.3 Members Receiving Feedback

We consider different subsets of meeting members to receive feedback, and discuss the member subsets that do not make members, especially those who are less active in their participation, feel negatively about feedback, and make members participate in a meeting. Meetings are essentially conversations. In most conversations, there is a turn-taking system, which operates in the following manner. In a conversation between two people, there is always one listener, who will always become the next speaker when the current speaker stops speaking. In a conversation between three or more people, the next speaker is selected based on the turn-taking rules (Sacks et al. (1974)) shown in Table 1. When a listener wants to speak next, he/she communicates his/her intention to speak through eye contact or a gesture to the current speaker, and he/she either is selected as the next speaker by the current speaker (rule (a)) or voluntarily starts speaking (rule (b)).

Table 1: *Turn-taking rules*

(a) Selection by others: The current speaker selects the next speaker.
(b) Selection by self: A next speaker self-selects him/herself, if (a) does not occur.
(c) Continuation: The current speaker may continue speaking, if neither (a) nor (b) occurs.

We now consider the following three scenarios, in which different members receive feedback, and examine which scenario makes less active members feel least negatively about feedback.

In the first scenario, a member whom other members would like to hear speak next, or a member who is willing to speak next (a potential speaker) receives feedback. Upon receiving feedback, a potential speaker voluntarily speaks. This means that a potential speaker takes action following either rule (a) or rule (b) in Table 1. In this scenario, a less active member may not feel negatively because other members are not aware that the member received feedback.

In the second scenario, a member who is currently speaking (a current speaker) receives feedback. Upon receiving feedback, the current speaker identifies a potential speaker among other members and encourages the identified potential speaker to speak. Namely, the current speaker follows rule (a). Identifying a potential speaker and encouraging to that member to speak are relatively straightforward, because the current speaker is likely to be a facilitator or leader. In this scenario, a less active member may not feel negatively because he/she

receives feedback, not directly from the computer, but indirectly through the current speaker, a peer member.

In the third scenario, all members receive feedback. Upon receiving feedback, each member identifies a potential speaker among all members. If a member identifies him/herself as a potential speaker, he/she voluntarily speaks. If a member identifies another member as a potential speaker, he/she encourages the identified potential speaker to speak. In this scenario, a less active member may not feel negatively. This is because, in the former case, who is the potential speaker is not explicitly disclosed to other members, and in the latter case, he/she receives feedback indirectly through the peer member.

5. FIELD STUDY

We designed a field study in brainstorming sessions under the following four feedback conditions. We compared Vibe-All and Light-All to answer RQ1, and compared Vibe-PS, Vibe-CS, and Vibe-All to answer RQ2.

Vibe-Potential Speaker (Vibe-PS): The system provides vibrotactile feedback only to a potential speaker. The feedback recipient is to voluntarily speak.

Vibe-Current Speaker (Vibe-CS): The system provides vibrotactile feedback only to the current speaker. The feedback recipient is to identify a potential speaker and encourage the identified potential speaker to speak.

Vibe-All: The system provides vibrotactile feedback to all members. Each feedback recipient is to identify a potential speaker. If a member identifies him/herself as a potential speaker, he/she is to voluntarily speak. If a member identifies another member as the potential speaker, he/she is to encourage the identified potential speaker to speak.

Light-All: The system provides visual feedback to all members. Each feedback recipient is to take the same action as the members in Vibe-All.

5.1 Study Design

We employed a within-subjects (i.e., within-groups) design and assigned participants to four mixed-gender groups, each with four or five participants. Each group performed brainstorming sessions in all four feedback conditions (Figure 1). The independent variable was the types of feedback forming four conditions. The dependent variables were shown in Figure 3. With the three Vibe conditions, the seat of each participant's chair vibrated (Figure 2c). With the Visual condition, the spotlight on the ceiling flashed (Figure 2b). To prevent participants from brainstorming on the same topic in subsequent conditions, four different topics were utilized. The feedback conditions were counterbalanced and the order of topics was randomized.

5.2 Setup

We partnered with a company and conducted a study in the field with brainstorming sessions that its employees hold as a part of their regular business, because it is easier to reproduce social behavior in the field than in a laboratory (Hornecker and Nicol (2012)). However, conducting a study in the field came with its own limitations: we were not able to control all variables such as the composition of groups, nor were we able to use a large number of groups or consider other feedback conditions.

All brainstorming sessions were held in the same meeting room (Figure 2a) in the company's building.

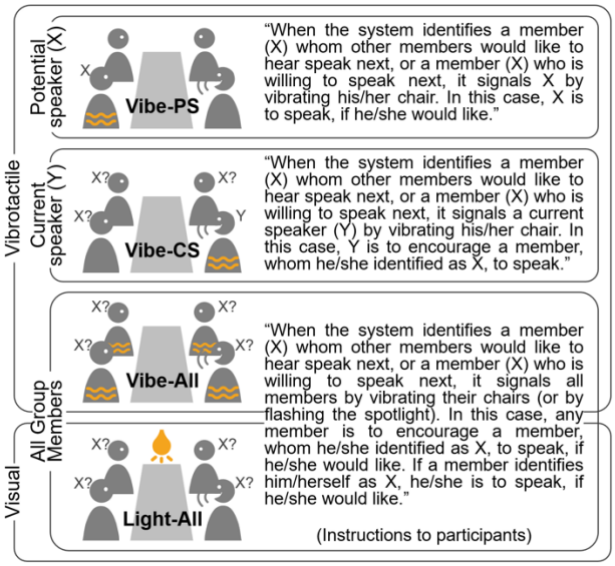


Figure 1: *Study conditions and instructions to participants*

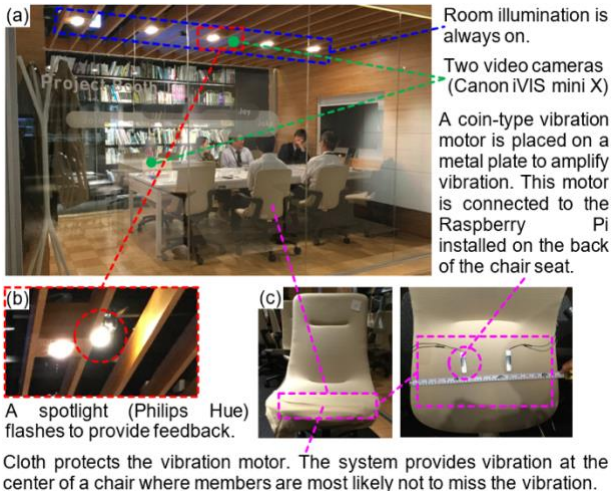


Figure 2: *Study environment*

Participants sat around a rectangular table in each session. Before starting our study, we confirmed that, when a participant received vibrotactile feedback, other participants sitting near him/her would not hear the vibration sound. During the

session, we recorded audio and video using two video cameras: one on the table that the participants used and the other on the ceiling of the room (Figure 2a).

5.3 Wizard of Oz Method to Provide Feedback

To fully answer RQ1 and RQ2, computers in meeting support must be highly intelligent and comparable to a human facilitator. However, such computers are beyond what is currently available. We therefore adopted a Wizard of Oz method.

During the session, the wizard observed participants who were in the meeting room, from a separate room. The wizard listened to live audio and watched live video from video cameras in real time and operated the feedback application we developed, to send feedback to participant(s). When the wizard touched the trigger button, the application sent a signal to the chair of the member(s) chosen to receive feedback or to the spotlight on the ceiling.

The wizard sent feedback (touched the trigger button) to meet the following requirement:

- When the wizard identifies a member who meets either (1) or (2) below during a session, she always sent feedback: (1) a member who has been less active in terms of verbal behavior and whom other members would like to hear speak next, or (2) a member who is willing to speak next and expresses his/her intention to speak through non-verbal behavior (for example, facial expressions, gesture, or posture).
- The wizard sent feedback at least once during each session to every member.

We hired a professional facilitator as the wizard, to help that the wizard properly identifies a member who meets above requirements.

5.4 Participants

Seventeen participants (12 males and 5 females, average age of 39, age 25–59) took part in the study. We assigned participants to four mixed-gender groups, each with four or five participants. All participants work in the company we partnered with and know each other. They engage in research and development and regularly hold brainstorming sessions.

5.5 Task and Procedure

Each group conducted four sessions, each lasting approximately 40 minutes. Each group was given four topics to discuss: ideas to support employees (1) to perform their individual work, (2) to perform their group work, (3) to conduct activities during their breaks, and (4) to perform their work outside the company premises, in a manner that increases their work productivity and enhances their healthy lifestyle. This company designs and manufactures office furniture and provides solutions to improve work environment in the office. The four topics above were not prepared for the sake of our study but were the actual topics that the company employees were to discuss as a part of their regular business at the time of our study.

Before each session, we instructed the participants how our system behaves and what the feedback means (Figure 1). Note that the participants knew which feedback conditions were used in the session. We also explained that, when participants receive feedback during the session, they could voluntarily decide whether and when to act on the feedback. During the session, the wizard monitored live audio and video from the camera and operated the feedback application. After each session, we asked participants to complete a 5-point Likert scale questionnaire. After all sessions were completed, we asked the participants to complete multiple choice questionnaire. We then conducted a semi-structured group interview with each group. We also conducted a test to examine each participant's personality (how active he/she is, see "Introversion-Extroversion Index (IEI)" in section 5.6) after all sessions were completed. In addition, we conducted an informal interview with the wizard. Finally, this study was approved by the research ethics committee of our institution.

5.6 Measures

To answer our research questions, we used quantitative and qualitative measures (Figure 3).

5.6.1. Probability of Turn-Taking following the Feedback

Using recorded videos, we examined whether turn-taking occurred following the feedback. Table 2 shows the criteria used to determine whether turn-taking occurred. Using these criteria, we obtained

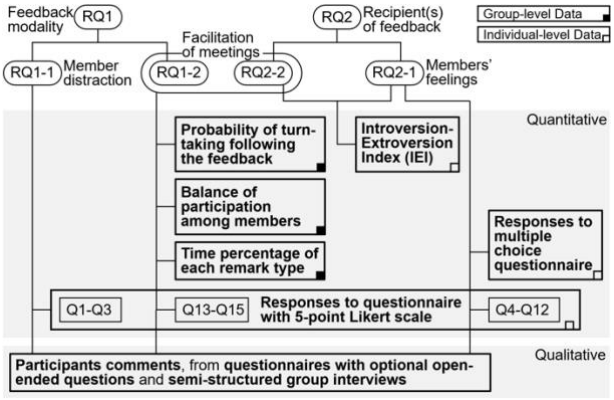


Figure 3: *Research questions and measures*

Table 2: *Criteria for turn-taking*

Feedback conditions	Feedback Recipient(s)	Who is speaking when feedback is sent	Criteria for turn-taking
Vibe-PS	X	Someone other than X is speaking.	X started speaking.
		No one is speaking (silence).	X started speaking.
Vibe-CS	Y	Y is speaking.	Someone other than Y started speaking.
Vibe-All, Light-All	All group members	Z is speaking.	Someone other than Z started speaking.
		No one is speaking (silence).	Someone started speaking.

("Speaking" in this table excludes back-channel feedback.)

Table 3: *Types of remarks*

Remarks type	Description
Idea	Suggesting an idea related to the meeting subject
Agreement	Expressing agreement or a supporting opinion
Disagreement	Expressing disagreement or an opposing opinion
Discussion	Discussing ideas
Facilitation	Commenting to help facilitate the meeting
Other	Making a remark outside of the main subject of the meeting

the probability with which the feedback from the wizard caused turn-taking to occur, defined as the ratio of the number of feedback events that led to turn-taking to the total number of feedback events during the session. We used three time intervals: 10, 20, and 30 seconds after the feedback.

5.6.2. Balance of Participation among Members

Previous work (e.g., DiMicco et al. (2007); Terken and Sturm (2010); Tausczik and Pennebaker (2013)) uses the modified Gini coefficient (Weisband et al. (1995)) to measure the balance of participation among members. The Gini coefficient is a measure of inequality, ranging from 0 (perfectly equal) to 1. The balance of participation is obtained by subtracting the Gini coefficient from 1.

Using recorded videos, we obtained the balance of participation regarding speaking activities, such as the total speaking length (total time length of all remarks of each member) and the speaking frequency (frequency of speaking of each member), similarly to the existing work.

5.6.3. Time Percentage of Each Remark Type

Using recorded videos, we examined what types of remarks members made during the sessions, in the following manner. Using a modified version of the remark types defined by Leshed et al. (Leshed et al. (2009)), we coded each remark from members into six types (Table 3). We ignored backchannel (a short acknowledgement) (Den et al. (2011)), laughter, and filler. Each remark is separated by a turn-taking between speakers. The coding of the remark type was done independently by two coders. Cronbach's alpha by coders was 0.85 and coders settled disagreements through discussion. For each type of remark given by all members in a session, we obtained the time percentage of the remark type, defined as the ratio of the total speaking length of the remark type to the total speaking length of all remark types combined.

5.6.4 . Questionnaire to Participants

5-point Likert Scale Questionnaire. The questionnaire distributed to the participants after each session contained 15 questions: three for distraction of participants, nine for members' positive or negative feeling, and three for facilitation of meetings. Each participant answered using the Likert scale of 5 levels (1: strongly disagree, 2: disagree, 3: neither agree nor disagree, 4: agree, and 5: strongly agree).

Multiple Choice Questionnaire. The questionnaire conducted after all sessions contained three questions: "which system do you like best?", "with which system do you feel most comfortable?", and "with which system are you most satisfied concerning the productivity of the brainstorming sessions?" For each question, each participant chose one of the four feedback conditions.

Questionnaires with Optional Open-ended Questions. The above two questionnaires also asked participants to provide free-form comments on the feedback that they experienced.

5.6.5. Semi-structured Group Interviews

To understand how participants experienced different feedback conditions, we conducted a semi-structured group interview with each group after all sessions.

5.5.6. Introversion-Extroversion Index (IEI)

To explore how members' personalities (whether they are more active or less active) impacts members' feeling about feedback (RQ2-1) and facilitation of meetings (RQ2-2), we conducted an Awaji-Okabe introversion/extroversion test (Awaji et al. (1932)) with each participant, to obtain his/her introversion-extroversion index (IEI). This test consists of 50 questions. We equated extroverted and introverted with more active and less active, respectively[5].

To address RQ2-1, we obtained the correlation of participants' IEIs with their ratings on Likert scale questionnaire (Q4–Q12) and with feedback condition selected in responses to the multiple

[5] We calculated Pearson's correlation coefficients between participants' IEIs and their total speaking length, and found positive correlation for all feedback conditions (Vibe-PS: r=0.73, p<0.01; Vibe-CS: r=0.74, p<0.01; Vibe-All: r=0.73, p<0.01; Light-All: r=0.68, p<0.01). We also found similar correlation coefficients between their IEIs and speaking frequency.

choice questionnaire. To address RQ2-2, we obtained correlation of participants' IEIs with their probabilities of turn-taking following the feedback and with their ratings on Likert scale questionnaire (Q13–Q15).

5.7 Data Analysis

We analyzed the data collected in our study. Data on feedback provided to participants owing to errors in the wizard's operation and judgment were excluded from our analysis.

Group-level Data (rectangular boxes with a filled square in Figure 3)**.** Regarding the probability of turn-taking following feedback (Figure 4), the balance of participation among members (Table 4), and the time percentage of each remark type (Table 5), we conducted a one-way repeated-measures (RM) ANOVA (within group factor: feedback conditions) with post hoc pairwise comparisons through using the Bonferroni adjustment. We analyzed the correlation between participants' IEIs and their probabilities of turn-taking following the feedback, using Pearson's correlation coefficients.

Individual-level Data (rectangular boxes with a square in Figure 3)**.** Regarding participants' responses to the 5-point Likert scale questionnaire (Table 6), we conducted a one-way repeated-measures (RM) ANOVA (within subject factor: feedback conditions) with post hoc pairwise comparisons through using the Bonferroni adjustment. We analyzed the correlation between participants' IEIs and the ratings they provided in the 5-point Likert scale questionnaire, using Pearson's correlation coefficients.

Regarding participants' responses to the ordinal scale multiple choice questionnaire (Figure 6), we used a $\chi 2$ test followed by a Ryan's multiple comparison test for proportions, because participants' responses are categorical data (data that relate the frequency of each target category to the total frequencies). We analyzed, using the correlation ratio, the correlation between participants' IEIs and feedback condition selected in responses to the multiple choice questionnaire.

Participants Comments. We conducted qualitative inductive analysis of the transcripts of the group interviews and written comments provided in the questionnaires with optional open-ended questions.

6. RESULTS

Sixteen brainstorming sessions were conducted. Each session lasted 40.9 minutes, on average. The wizard provided feedback 21.9 times per session, on average. Each participant received feedback in a session the following number of times, on average: Vibe-PS: 6.1; Vibe-CS: 4.6; Vibe-All: 21.5; Light-All: 20.3. Error bars in the figures in this section show the standard error of the mean. Asterisks in figures and tables show significance differences (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$) in the post hoc test described in section 5.7.

6.1 Probability of Turn-Taking following the Feedback

Figure 4 shows the probability of turn-taking following the feedback at three different time intervals (10, 20, and 30 seconds after the feedback). One-way RM ANOVA on the probability at these time intervals showed a significant main effect of feedback condition at all time instances (10 seconds: $F(3,9) = 22.178$, $p = 1.71E-04$; 20 seconds: $F(3,9) = 10.00$, $p = 0.003$; 30 seconds: $F(3,9) = 7.983$, $p = 0.007$). Post hoc tests revealed significant differences between the four feedback conditions, as shown in Figure 4. For all three time intervals, the probability for Vibe-PS was significantly lower than that for all other feedback conditions. At 10 seconds, the probability for Vibe-All was significantly higher than that for Light-All. Only one member receives feedback with Vibe-PS and Vibe-CS, whereas all (four or five) members receive feedback with Vibe-All and Light-All. Therefore, the probability of turn-taking following feedback would seem to be much larger with Vibe-All and Light-All than with Vibe-PS and Vibe-CS. However, our results differed from this intuition.

Pearson's correlation coefficients between participants' IEIs and their probabilities of turn-taking were not significant for Vibe-PS, Vibe-CS, or Vibe-All.

6.2 Balance of participation among Members

Table 4 shows the balance of participation among the members in both total speaking length and speaking frequency. One-way RM ANOVA on the balance of participation for both did not show a significant main effect of feedback condition.

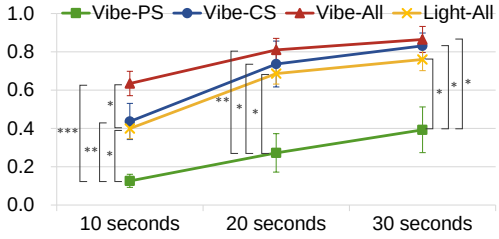


Figure 4: *Probability of turn-taking following the feedback*

Table 4: *Balance of participation among members*

	Vibe-PS	Vibe-CS	Vibe-All	Light-All
Total Speaking Length	0.71	0.74	0.69	0.68
Speaking Frequency	0.76	0.75	0.73	0.73

Table 5: *Time percentage of each remark type*

	Vibe-PS	Vibe-CS	Vibe-All	Light-All
Idea	28.1	28.0	31.2	30.2
Agreement	11.3	13.4	9.8	10.6
Disagreement	7.2	6.5	8.0	4.7
Discussion	33.2	35.4	37.0	34.2
Facilitation	4.1	4.3	4.0	7.7
Other	16.1	12.4	10.0	12.6
Total	100.0	100.0	100.0	100.0

6.3 Time Percentage of Each Remark Type

Table 5 shows the time percentage for each remark type, the ratio of the total speaking length of the remark type to the total speaking length of all remark types combined. One-way RM ANOVA on the time percentage for each remark type did not show any significant main effect of feedback condition for any remark type.

6.4 Questionnaire to Participants

5-point Likert Scale Questionnaire. Table 6 shows mean and one-way RM ANOVA results on the ratings of the 5-point Likert scale questionnaire. Participants agreed that Vibe-All distracted more participants than Light-All (Q2); Vibe-PS made it easier to understand the intent of the system than Vibe-CS (Q5); Light-All produced discussion results with which participants strongly agreed, and/or were strongly satisfied, than Vibe-CS (Q9); and Vibe-PS and Vibe-CS resulted in more diverse opinions from members than Light-All (Q15). With respect to Q4–Q12 and Q13–Q15, Pearson's correlation coefficients between participants' IEIs and their ratings were not significant for Vibe-PS, Vibe-CS, or Vibe-All.

Multiple Choice Questionnaire. Figure 6 shows the frequency rate of each of the four feedback conditions in the multiple choice questionnaire. Participants chose Vibe- CS least in all three questions. A $\chi 2$ test for each question showed a significant main effect of feedback condition for comfortability, $\chi 2(3)=8.00$, $p=0.046$. A post hoc Ryan test revealed that the value of comfortability

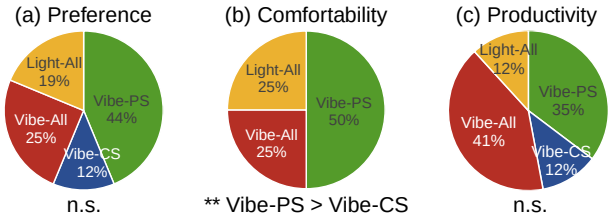


Figure 6: *Responses to multiple choice questionnaire.*

for Vibe-PS was significantly higher than for Vibe-CS, $p=0.008$ (Figure 6b). For all three questions, the correlation ratio between participants' IEIs and feedback condition selected by participants was not significant.

6.5 Participants Comments

We conducted qualitative inductive analysis of participants comments and the following themes were identified: encouraging actions, sense of experience sharing, sense of participation, identifying a potential speaker, and sense of comfort. Sense of experience sharing refers to the sense of involvement that a member develops from sharing the same goals and values with other members of the group. Sense of participation refers to the sense of self-involvement that a member develops from addressing an issue of the group together with other members.

6.5.1. Feedback Modality (RQ1)

Encouraging Actions. Our results suggest that vibrotactile feedback encouraged members to take voluntary actions. In contrast, visual feedback using light did not.

> [Vibration (vibrotactile modality)]
> P11: Vibration motivated me to help advance meetings.
> P15: I felt encouraged to listen and speak.
> P13: Vibration motivated me to speak.
>
> [Light (visual modality)]
> P01: I felt less pressure with light than with vibration, and it led me to often ignore the light feedback.
> P02: Light did not encourage me to take an action.
> P08: Light made me feel that it was for someone else.

Sense of Experience Sharing and Sense of Participation. Our results suggest that visual feedback using light helped to create a strong sense of experience sharing and a weak sense of participation. In contrast, the vibrotactile feedback led to a weak sense of experience sharing and a strong sense of participation. During a group interview, one participant commented on the difference that vibration and light created in his

Table 6: *Responses to 5-point Likert scale questionnaire*

RQ			Questions	Vibe -PS	Vibe -CS	Vibe -All	Light -All	ANOVA F(3,45)	ANOVA p-value		Bonferroni's post-hoc	Bonferroni's post-hoc p-value	
RQ1-1 Member distraction		Q1	I immediately noticed receiving feedback from the system.	4.50	4.31	4.69	3.94	2.284	0.092				
		Q2	I was distracted by the feedback from the system.	3.38	3.00	3.63	2.50	3.108	0.036	*	Vibe-All > Light-All	0.047	*
		Q3	I constantly anticipated feedback from the system and could not concentrate on the discussion.	2.25	2.13	2.00	1.75	2.075	0.117				
RQ2-1 Members' feelings	Feelings about overall usefulness	Q4	I would like to use this system at other meetings.	2.56	2.75	2.94	3.00	2.143	0.108				
		Q5	It was easy to understand the intent of the system.	4.44	3.63	4.13	3.81	3.025	0.039	*	Vibe-PS > Vibe-CS	0.043	*
		Q6	I thought that the system was effective overall.	3.06	2.81	2.88	2.69	1.398	0.256				
	Sense of achieve-ment and satisfaction	Q7	I am satisfied with the result (quality) of the discussion.	3.56	3.25	3.56	3.56	0.531	0.663				
		Q8	I am satisfied with the result (quantity) of the discussion.	3.69	3.13	3.63	3.81	2.778	0.052				
		Q9	I agree with/I am satisfied with the result of the discussion.	3.94	3.44	3.88	4.19	3.987	0.013	*	Light-All > Vibe-CS	0.019	**
	Frustration	Q10	I felt pressure from the system.	2.81	2.69	2.44	2.38	0.968	0.416				
		Q11	I felt that the system was too intrusive and interfering.	3.06	2.56	2.94	2.31	2.682	0.058				
		Q12	I felt that the system was mysterious and sometimes scary.	2.56	2.19	2.38	2.19	1.983	0.130				
RQ1-2, RQ2-2 Facilitation of meetings		Q13	The system equally encouraged members to participate/speak.	3.38	3.00	3.00	3.00	1.246	0.304				
		Q14	The system encouraged collaborative participation of members.	3.38	3.13	3.31	2.94	2.064	0.118				
		Q15	The system extracted diverse opinions from members.	3.19	3.00	2.88	2.44	5.792	0.002	**	Vibe-PS > Light-All	0.034	*
											Vibe-CS > Light-All	0.045	*

sense of experience sharing and sense of participation (see below), and other participants largely agreed with his statement.

> [Vibration (vibrotactile modality) vs. Light (visual modality)]
> P05: Light created a strong sense of feedback sharing. As it is obvious that all members received the light, I often felt "I do not need to take an action, because others also received the light and may take an action" and became dependent on others. On the other hand, although I knew that all members received the vibration, vibration is not visible and did not create a strong sense of feedback sharing. As a result, I hardly became dependent on others.

6.5.2. Members Receiving Feedback (RQ2)

Sense of Participation. Our results suggest that a potential speaker receiving feedback (Vibe-PS) helped to create a weak sense of participation. In contrast, either a current speaker (Vibe-CS) or all members (Vibe-All and Light-All) receiving feedback did not appear to have the same effect.

> [Vibe-PS]
> P16: Even if I do not speak, others will not notice and will not think "that member has not spoken".
> P05: Even if I do not speak, it does not hinder the meeting.

Identifying a Potential Speaker. Vibe-CS, Vibe-All, and Light-All expect each member to identify a potential speaker. Our results suggest that, when the current speaker received feedback (Vibe-CS), he/she found it difficult to identify a potential speaker. In contrast, when all members received feedback (Vibe-All and Light-All), members felt positively about flexibility of identifying a potential speaker.

> [Vibe-CS]
> P10: When I received a message from the system, it was hard to determine to whom I should encourage to speak.
> P09: I found it difficult to think about both what I was talking about and to whom I should encourage to speak.
> P06: This (Vibe-CS) may be suitable for someone with a certain level of meeting facilitating skills.
>
> [Vibe-All, Light-All]
> P08: Ambiguous message (e.g., "this feedback may be for me or may be for someone else") enabled me to actively participate.
> P17: I felt that the system was flexible and allowed me to act based on my own will.

Sense of Experience Sharing. Our results suggest that all members receiving feedback (Vibe-All and Light-All) helped to create a strong sense of experience sharing and, as a result, members felt that it was easier to take voluntary actions.

> [Vibe-All, Light-All]
> P03: I believe that all members receiving feedback contributed to the sense of experience sharing and made everyone feel that he/she should participate and advance the meeting.
> P10: Because everyone knows, it was easier to proceed with the conversation.
> P15: I felt that all members being aware raised the level of recognition.
> P03: I felt that it was easier to speak, because there was no pressure from being the only one who received feedback.

7. DISCUSSION

7.1 Feedback Modality (RQ1)

7.1.1. Distraction of Members (RQ1–1)

Although members felt significantly more distracted by vibrotactile feedback (Vibe-All) than visual feedback (Light-All), their ratings were not high for either (Table 6, Q2). In addition, for both types of feedback, members felt that they could mostly concentrate on the discussion even with feedback (Table 6, Q3). Furthermore, members sometimes ignored the visual feedback (Section 6.5.1). Considering that feedback should neither distract members nor be ignored by members, we conclude that vibrotactile feedback provides a moderate level of distraction compared to visual feedback.

7.1.2. Facilitation of Meetings (RQ1–2)

Vibrotactile feedback (Vibe-All) encouraged significantly more active turn-taking than visual feedback (Light-All) immediately after feedback (Figure 4). This is because vibration encouraged members to take actions (Section 6.5.1) and helped them to organize more active turn-taking.

In addition, vibrotactile feedback (Vibe-All) helped to create a weak sense of experience sharing and a strong sense of participation and, in contrast, visual feedback (Light-All) helped to create a strong sense of experience sharing and a weak sense of participation (Section 6.5.1). The social compensation effect (Williams and Karau (1991)) and social loafing effect (Latané et al. (1979)) may explain these findings with vibrotactile and visual feedback, respectively.

7.2 Members Receiving Feedback (RQ2)

7.2.1. Feeling of Members (RQ2–1)

When members received feedback as a potential speaker (Vibe-PS) rather than a current speaker (Vibe-CS), they felt significantly more positive, i.e., easier to understand the intent of the system (Table 6, Q5) and more comfortable (Figure 6b). This is because, with Vibe-PS, it is clear what the feedback recipient (the potential speaker) is expected to do: to speak voluntarily. In contrast, with Vibe-CS, the feedback recipient (the current speaker) is expected to identify a potential speaker and to encourage that member to speak. The current speaker who received feedback often found this difficult (Section 6.5.2). As a result, they felt it more difficult to understand the intent of the system and became less comfortable, so they felt more negatively about feedback.

Vibe-CS and Vibe-All both require members to identify a potential speaker, but our results show that members responded to them differently: there were no comments indicating difficulty in identifying a potential speaker with Vibe-All, unlike with Vibe-CS; there were even some positive comments appreciating the flexibility in identifying the potential speaker with Vibe-All. This is because, with Vibe-All, all members received feedback. This would lead to a bystander

effect (Latané and Darley (1970)), and increased members' tolerance for requests from the system.

We now discuss how participants' personality (whether they are more active or less active) affected participants' feelings about feedback. As discussed in section 2, with existing meeting support systems where all members receive feedback, less active members often feel negatively about feedback. On the contrary, in our study, participants' personality (IEI) did not impact the questionnaire responses for any of the three subsets of members receiving feedback (Vibe-PS, Vibe-CS, and Vibe-All) (Section 6.4).

7.2.2. Facilitation of Meetings (RQ2–2)

Either all members (Vibe-All) or the current speaker (Vibe-CS) receiving feedback resulted in significantly more active turn-taking than the potential speaker receiving feedback (Vibe-PS) (Figure 4) This is because, with Vibe-All, members tended to develop a strong sense of experience sharing, thereby making it easier for individual members to take action (Section 6.5.2) and members might increase their efforts not due to the social loafing effect (Latané et al. (1979)) but rather due to the social compensation effect (Williams and Karau (1991)). As a result, there was significantly more active turn-taking than with Vibe-PS. In addition, this could be explained by social facilitation[6] (Triplett(1898); Allport (1924)), although it is not observed in participants comments. Note that one member receives feedback when both the current speaker (Vibe-CS) and the potential speaker (Vibe-PS) receive feedback. Upon receiving feedback, the potential speaker (Vibe-PS) developed a weaker sense of participation than the current speaker (Vibe-CS) (Section 6.5.2). As a result, with Vibe-CS, there was significantly more active turn-taking than with Vibe-PS.

We now discuss how participants' personality affected facilitation of meetings. As discussed in section 2, in some existing meeting support systems, less active members did not necessarily increase their participation, whereas more active members usually decreased their participation. On the contrary, in our study, participants' personality (IEI) had no impact on the probability of turn-taking following feedback in any of the three subsets of members receiving feedback (Vibe-PS, Vibe-CS, and Vibe-All) (Section 6.1).

8. LIMITATIONS

We now discuss some limitations of our study. First, our study did not include the baseline of the comparison: that is, no studies were conducted without feedback. Feedback conditions we used may or may not be more effective than the baseline in supporting meetings.

Second, in comparing different modalities (Vibe-All and Light-All), each member received feedback "privately" with Vibe-All, and all members shared the same light feedback and received feedback "publicly" with Light-All. These "private and public" aspects of Vibe-All and Light-All may or may not have contributed to our findings, and it is not clear how much they contributed to the difference in the sense of experience sharing. When members "privately" receive visual feedback—for instance, through their smart phones—our findings may need to be revised.

Third, our sample size is small. A larger sample size may change our results, such as the balance of participation, the time percentage for each remark type, and the correlation between participants' IEIs and some data, where there were no statistically significant between the feedback conditions.

9. CONCLUSION

We conducted a field study exploring how members modify their responses to feedback, when different modalities are used for feedback and when different subsets of members receive feedback. Table 7 and Table 8 summarize our findings on two key research questions. We hope our findings inspire designers, developers, and researchers for meeting support and other types of group collaboration support as well as vibrotactile interfaces.

Table 7: *Answers to RQ1 related to feedback modality*

		Vibe-All	Light-All
RQ1-1	Member distraction	Moderate level	Low level enough to be ignored
RQ1-2	Facilitation of meetings	More likely to cause turn-taking	Less likely to cause turn-taking

Table 8: *Answers to RQ2 related to feedback recipients*

		Vibe-PS	Vibe-CS	Vibe-All
RQ2-1	Members' feelings	Positive	Negative	In between Vibe-PS and Vibe-CS
RQ2-2	Facilitation of meetings	Less likely to cause turn-taking	More likely to cause turn-taking	More likely to cause turn-taking

10. REFERENCES

Allport, F.H. (1924) *Social psychology*. Reverside Press, Cambridge, MA, USA.

Awaji, E. and Okabe, Y. (1932) The Introversion-Extroversion Test. *The Japanese Journal of Psychology*, 7, 1, 2–10.

Balaam, M., Fitzpatrick, G., Good, J. and Harris, E. (2011) Enhancing interactional synchrony with an ambient display. In *Proceedings of the SIGCHI*

[6] Social facilitation is a phenomenon whereby increased task performance comes about by the mere presence of others, who do the same task (co-action effect) or who is a passive spectator/audience (audience effect).

Conference on Human Factors in Computing Systems (CHI '11). ACM, NewYork, NY, USA, 867–876.

Bachour, K., Kaplan, F. and Dillenbourg, P. (2010) An interactive table for supporting participation balance in face-to-face collaborative learning, *IEEE Transactions on Learning Technologies* 3, 3, 203-213.

Bergstrom, T and Karahalios, K. (2007) Seeing more: Visualizing audio cues. In *Proceedings of the 11th IFIP TC.13 International Conference on Human-Computer Interaction* (INTERACT 2007), Springer, 29–42.

Birnholtz, J. Gergle, D., Liebman, N. and Sinclair, S. (2015) Feeling Aware: Investigating the Use of a Mobile Variable-Friction Tactile Display for Awareness Information. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI '15). ACM, NewYork, NY, USA, 16–25.

Blum, J.R. and Cooperstock, J.R. (2016) Expressing Human State via Parameterized Haptic Feedback for Mobile Remote Implicit Communication. In *Proceedings of the 7th Augmented Human International Conference 2016* (AH '16). ACM, NewYork, NY, USA, 41.

Burns, G. (1995) The secrets of team facilitation. Training and Development 49, 6, 46-52.

Dagan, E., Segura, E.M., Flores, M. and Isbister, K. (2018) “Not Too Much, Not Too Little” Wearables For Group Discussions. In *Proceeding CHI EA '18 Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI EA ’18). ACM, NewYork, NY, USA, LBW129.

Dagan, E., Segura, E.M., Bertran, F.A., Flores, M., Mitchell, R. and Isbister, K. (2019) Design Framework for Social Wearables Proceeding. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (DIS '19). ACM, NewYork, NY, USA, 1001–1015.

Den, Y., Yoshida, N., Takanashi, K. and Koiso, H. (2011) Annotation of japanese response tokens and preliminary analysis on their distribution in three-party conversations. In *Proceedings of the 14th Oriental COCOSDA* (O-COCOSDA 2011). IEEE, 168–173.

DiMicco, J.M., Pandolfo, A. and Bender, W. (2004) Influencing group participation with a shared display. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (CSCW ’04). ACM, NewYork, NY, USA, 614–623.

DiMicco, J.M., Hollenbach, K.J., Pandolfo, A. and Bender, W. (2007) *The Impact of Increased Awareness While Face-to-Face. Human-Computer Interaction*, 22, 47–96.

Doyle, M. and Straus,D. (1993) *How to make meetings work: The new interaction method*. Berkley Trade, NewYork, NY, USA.

Hornecker, H. and Nicol, E. (2012) What do lab-based user studies tell us about in-the-wild behavior? : insights from a study of museum interactives. In *Proceedings of Designing Interactive Systems* (DIS '12). ACM, New York, USA, 358–367.

Kim, T., Chang, A., Holland, L. and Pentland, A.S. (2008) Meeting mediator: enhancing group collaboration using sociometric feedback. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (CSCW '08). ACM, New York, NY, 457–466.

Latané, B. and Darley, J.M. (1970) The unresponsive bystander: Why doesn't he help? Appleton-Century-Crofts, New York.

Latané, B., Williams, K. and Harkins, S. (1979) Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology* 37, 6, 822–832.

Leshed, G., Perez, D., Hancock, J.T., Cosley, D., Birnholtz, J., Lee, S., McLeod, P.L. and Gay, G. (2009) Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09). ACM, NewYork, NY, USA, 537–546.

Meier, A., Matthies, D.J.C., Urban, B. and Wettach, R. (2015) Exploring vibrotactile feedback on the body and foot for the purpose of pedestrian navigation. In *Proceedings of the 2nd international Workshop on Sensor-based Activity Recognition and Interaction* (iWOAR '15). ACM, NewYork, NY, USA, 11.

Nowak, M., Kim, J., Kim, N.W. and Nass, C. (2012) Social visualization and negotiation: effects of feedback configuration and status. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (CSCW '12). ACM, NewYork, NY, USA, 1081–1090.

Occhialini, V., Essen, H.v. and Eggen, B. (2011) Design and Evaluation of an Ambient Display to Support Time Management during Meetings. In *Proceedings of the 13th IFIP TC 13 international Conference on Human-computer interaction* (INTERACT '11). Springer-Verlag Berlin, Heidelberg, 263-280.

Oxley, N.L., Dzindolet, M.T. and Paulus, P.B. (1996) The effects of facilitators on the performance of brainstorming groups. *Journal of Social Behavior and Personality*, 11, 4, 633–646. Select Press, Corte Madera, CA, USA.

Pielot, M. and Oliveira, R. (2013) Peripheral Vibro-Tactile Displays. In *Proceedings of the 15th international Conference on Human-computer interaction with mobile devices and services* (MobileHCI '13). ACM, NewYork, NY, USA, 1–10.

Reeves, B. and Nass, C. (1996) *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Chicago, IL, US: Center for the Study of Language and Information, New York, NY, US: Cambridge University Press.

Sacks, H., Schegloff, E.A. and Jefferson, G. (1974) A simplest systematics for the organization of turn-taking in conversation. Language 50, 4, 696–735.

Saket, B., Prasojo, C., Huang, Y.and Zhao, S. (2013) Designing an Effective Vibration-Based Notification Interface for Mobile Phones. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (CSCW '13). ACM, NewYork, NY, USA, 1499–1504.

Sanchez-Cortes, D., Aran, O., Mast, M.S. and Gatica-Perez, D. (2012) A Nonverbal Behavior Approach to Identify Emergent Leaders in Small Groups. *IEEE Transactions on Multimedia* 14, 3, 816–832.

Schiavo, G., Cappelletti, A., Mencarini, E., Stock, O. and Zancanaro, M. (2014) Overt or subtle? Supporting group conversations with automatically targeted directives. In *Proceedings of the 19th international Conference on Intelligent User Interfaces* (IUI '14). ACM, NewYork, NY, USA, 225–234.

Schneider, O.S., Seifi, H., Kashani, S., Chun, M. and MacLean, K.E. (2016) HapTurk Crowdsourcing Affective Ratings of Vibrotactile Icons. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, NewYork, NY, USA, 3248–3260.

Snyder, J., Matthews, M., Chien, J.T., Chang, P.F., Sun, E., Abdullah, S. and Gay, G. (2015) MoodLight: Exploring Personal and Social Implications of Ambient Display of Biosensor Data. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '15). ACM, NewYork, NY, USA, 143–153.

Soller, A., Martínez, A., Jermann, P. and Muehlenbrock, M. (2005) From Mirroring to Guiding: A Review of State of the Art Technology for Supporting Collaborative Learning. *International Journal of Artificial Intelligence in Education*, 15, 4, 261–290.

Streng, S., Stegmann, K., Hußmann, H. and Fischer, F. (2009) Metaphor or diagram? Comparing different representations for group mirrors. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group* (OZCHI '09). ACM, NewYork, NY, USA, 249–256. http://doi.org/10.1145/1738826.1738866

Tausch, S., Hausen, D., Kosan, I., Raltchev, A. and Hussmann,H. (2014) Groupgarden: Supporting Brainstorming through a Metaphorical Group Mirror on Table or Wall. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction* (NordiCHI '14). ACM, NewYork, NY, USA, 26–30.

Sarah Tausch, Stephanie Ta and Heinrich Hussmann. (2016) A Comparison of Cooperative and Competitive Visualizations for Co-located Collaboration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, NewYork, NY, USA, 5034–5039.

Tausczik, Y.R. and Pennebaker, J.W. (2013) Improving teamwork using real-time language feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, NewYork, NY, USA, 459–468.

Terken, J. and Sturm, J. (2010) Multimodal Support for Social Dynamics in Co-located Meetings. *Personal Ubiquitous Computing* 14, 8. Springer-Verlag London, UK, 703–714.

Triplett, N. (1898) The dynamogenic factors in pacemaking and Competition. *American Journal of Psychology* 9, 507–533.

Weisband, S.P., Schneider, S.K. and Connolly, T. (1995) Computer-Mediated Communication and Social Information: Status Salience and Status Differences. *Academy of Management Journal* 38, 4, 1124–1151.

Williams, K.D. and Karau, S.J. (1991) Social loafing and social compensation: The effects of expectations of co-worker performance. *Journal of Personality and Social Psychology* 61, 4, 570–581.

“He’s adorable and I want to take him home”. Trust Perceptions Before and After First-Time Encounters with Social Robots

Emilia Sobolewska
School of Computing
Edinburgh Napier University
e.sobolewska@napier.ac.uk

Nicholas Vanderschantz
Department of Computer Science
University of Waikato
vtwoz@waikato.ac.nz

Francesca Bitcon
School of Computing
Edinburgh Napier University
40297473@live.napier.ac.uk

For users who have not interacted with a robot their perceptions are shaped by society, inculturation, and popular media. With the steady increase of robots in workplaces, during consumer encounters, and increasingly in the home, research is required to explore user perceptions of human-robot interaction and trust. This position paper aims to investigate trust between adults who have not previously interacted with a social robot, and Softbank Robotics’ Nao, before, during, and after an encounter. A novel insight from our study revealed that a negative portrait of robots does not decrease trust levels.

Human-Robot Interaction, Human-Robot Trust, Social Robots, Trust Perception Score, Godspeed Questionnaire

1. INTRODUCTION

For many years, the depictions of robots in popular culture have been viewed as fantasies in the realm of science fiction. As technology has developed, these robots and robot-like agents are beginning to move from the pages and screens into our everyday lives. Given this, human-robot interaction studies have often considered user perceptions of trust. To date, exposure to robots in homes, workplaces, and the community is limited, and as such, acceptance and trust in robot agents may not be commonplace.

We argue that most human-robot interaction work is presently situated in the industrial workplace, or with specific sectors of vulnerable society, and therefore work is required in other sectors.

We invited adult participants to interact with a social robot for the first time and their preconceived notions about robots were analysed. Our study contributes to the growing field of human-robot trust with a specific focus on trust perceptions, before, during, and after an adult user's first-time encounters with a social robot.

2. RELATED WORK

As robots have been introduced into the industrial environment case studies of robot acceptance and trust in the workforce [1][2] and recommendations for the development of cobots [3] have been proposed. We seek to investigate user perception and trust of *social robots* which have become increasingly available commercially for home, school, and other non-industrial use.

Trust has been widely investigated in a range of contexts over many years, including interpersonal trust [4], human trust in technologies and autonomous systems [5], and human-robot trust [6]. Research has shown that with increased exposure to robots, perceived initial novelty diminishes, and preconceived ideas about robots' behaviours and capabilities can change and evolve [7]. It is considered that there are three categorisations of a person's propensity to trust a system—dispositional, situational, and learned trust [8][9]. Dispositional trust relates to a user's propensity to trust autonomous systems independent of the situation. Situational trust depends on the environment in which the system is being utilised and the users' own self-confidence in their abilities to operate it. Finally, learned trust is developed through a user's evaluations of the system with direct reference to their previous and current experience. Furthermore, research defines three types of human-technology trust [10]; functionality, helpfulness, and reliability. Functionality is the perception that the technology has the features to accomplish a specific task. Helpfulness, that technology provides adequate and responsive aid to the user. Reliability, that technology continues to operate properly and without error.

Although researchers have also looked into the association between moods and emotions and how they relate to trust [11], interestingly, to date, it has been common for human-robot studies with autonomous systems, cobots, and social robots to be undertaken at a "safe" distance with participants only observing a robot undertaking tasks [7][12]. This continued research into human-robot interaction without any direct contact has made it challenging to fully understand how and why trust is established. There is little evidence to show that the trust a human believes they will have when interacting with a robot will actually exist in a real-world setting.

Humanoid robots have been designed to be human-like in appearance. They usually have a head, torso, two arms, and two legs though some only have body elements from the waist up. This includes social robots pertinent to our investigations. One such humanoid robot is Nao produced by Softbank Robotics. Nao is a bipedal robot with an anthropomorphic design. It has been used in many studies with children and adults of all ages [13][14][15][16].

Anthropomorphism and anthropomorphic design are often used as interchangeable descriptors of robots though they have very different meanings. Anthropomorphic design in robots refers to the physical imitation of the human form [17] rather than the projection of characteristics onto a non-human agent. Anthropomorphism is the tendency of a person to attribute human qualities or behaviours to inanimate objects, particularly common with humanoid robots [18]. Engagement with social robots can be improved if they are deliberately designed to exhibit human social characteristics [17]. However, the research into the impact of the visual appearance of 2D robotic heads with three levels of human likeness revealed that participants did not find the more human-like robots in the study to be more compelling [19]. Further studies with people interacting directly with robots may help to clarify how and why anthropomorphism in interactions could influence the trust a human is willing to place in a robot.

3. METHOD

Our study consisted of three phases; a pre-intervention interview and questionnaire; a video & audio recorded interaction with a Softbank Robotics' Nao; followed by a post-intervention interview & a questionnaire. Interviews were audio-recorded while questionnaires were filled in by hand by the participant, and later manual transcription and analysis were completed by the researcher.

Participants were drafted using a snowball recruitment method. All of them had a working knowledge of the English language, were adults of 18 years or older, and were not part of a vulnerable group as detailed by the University Ethics requirements. Each participant met with the same researcher in a one-to-one session at the University. Overall, 9 participants (4 female/5 male) took part. The youngest participant was

18 years old and the oldest 45 years old with the mean age being 30.88 years old (SD 9.5). The majority of the participants identified as having British nationality (6/9), with one Australian, one North American, and one Polish participant. One person completed high school education; the remaining participants have undergraduate (3/9) and postgraduate (5/9) qualifications. Five out of nine participants were currently enrolled in postgraduate computing studies at Edinburgh Napier University.

Pre-Intervention procedures included consent and demographic data collection. The participants then completed the Trust Perception Scale (TPS) [20] and Godspeed questionnaires [21] as well as the semi-structured interviews, to investigate their prior knowledge and assumptions regarding robots. We recorded a base level for human-robot trust and potential anthropomorphism the participants may experience, based on their dispositional and learned trust influences [8].

The Observation phase began with the robot having been programmed to engage with the participant by introducing itself and asking a series of questions requiring yes or no answers. Each interaction lasted approximately three minutes and was intended to give an impression of what the robot can and cannot do. The robot stood and gestured during the encounter. At the end, Nao thanked the participant for their involvement and said goodbye before sitting down and "going to sleep". Participants' facial expressions were recorded using Nao's in-built cameras. The videos were uploaded into the iMotions software to be analysed by the Affectiva facial analysis engine to investigate the situational trust [8].

Post-Intervention questionnaires and interviews investigated participants' experiences with Nao and whether their perceptions or assumptions regarding robots had changed. Participants were again asked to complete both the TPS and Goodspeed Questionnaires. The tools and procedures were exactly the same as for the pre-intervention phase.

4. RESULTS

Here we describe the results of our pre- and post-intervention questionnaires and interviews as well as the emotion data captured using video footage.

The completed Trust Perception questionnaires were compiled and each participant's scores before and after their interaction with Nao were calculated [20]. Given space constraints, we report only the median TPS score before (41.67) and after (49.58) interaction for the entire participant sample. It was seen that with the exception of P5, the trust scores remained either stable (within 8% movement) or increased after the interaction with the robot. This indicates that that TPS scores did increase after interaction with Nao.

The completed pre- and post-interaction Godspeed questionnaires were compiled and each participant's perceptions of the robot's anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety was calculated [21] (Table 1).

Table 1: *Godspeed Questionnaire before and after human-robot intervention*

Semantic Differential Themes (out of 5)	Before (Median)	After (Median)
Anthropomorphism	2	3
Animacy	2.5	4
Likability	3	5
Perceived Intelligence	4	4
Perceived Safety	4	4

The results of the semantic differentials showed increased perceptions of Anthropomorphism, Animacy, and Likability after interaction with Nao. Perceived Intelligence and Safety were not affected by the interaction with Nao and remained positive.

During Pre-Human-Robot Interaction Interviews the participants were asked in broad terms, regarding their previous experiences with the robots. The results were quite uniform—four participants admitted to never having had an interaction with a robot, whilst the remaining five listed digital assistants (i.e., Apple Siri, Amazon Alexa), as their only experience. It was mentioned that irritation arises when these systems do not respond as users would expect them to. P3 described this, stating "I just want things to work. When they do not work, it's frustrating". The participants were asked whether they have familiarity with robots emerging from other sources. Some participants described watching the news, and videos from known corporations (Honda, Softbank, Boston Dynamics), robots in hospitality, care, and factories (P4, P6, P7). All of the participants were aware of the pop-cultural representation of robots in movies and literature; with a majority (5/9) pointing out, that the depiction of automatons is often negative, with two of them (P1, P6), referring to dystopian fiction.

> Obviously seen lots of dystopian films. It is interesting, it is much easier to remember the bad robots than good robots, and I am just instinctively thinking of**.** [P1]

Only P5, P7, P9 reported no negative preconceptions of robots, based on popular culture, and recognising that their pre-interaction image of robots is manipulated and distorted.

> I think some of the fictional portrayals are no doubt extremely misleading. So, things like that (…), the horror stories of the Matrix, and stuff like that. You can imagine that it gets people the heebie-jeebies, but (…) nothing that we've got at this stage is anywhere near that advanced. [P3]

When asked whether they felt robots would have a use in our schools, homes, or workplaces the answers were focused on robots being able to perform simple tasks. Two of the participants (P3, P7) were more apprehensive about the intentions of the human programmers than the robot itself.

> To be honest, I'd be more concerned about the people doing the programming and their intent because with robots, being computerised systems, they essentially do what they're programmed to do... [P3]

The Observation used a combination of video recordings analysed with Affectiva Engine and researcher observations to provide cursory emotion and trust results. Due to the camera placement, both Nao's and participants' movements resulted in incomplete data for automated facial analysis. Nonetheless, Affectiva findings suggest that participants had neutral emotions for a significant portion of the interaction (mean of 70% of the time). Positive emotions were recorded on average 6.35% and negative emotions were recorded on average 0.82% of the time. While this data is of a limited sample there was very little negative emotion experienced by participants.

Instead, using video footage we manually coded observable physical or verbal cues at six key points during the study. When Nao was unveiled and stood for the first time, we observed interest (2 participants), amusement (4), and uncertainty (2), with one participant openly showing joy and excitement. When Nao introduced itself, six participants smiled and showed an interest, one laughed, one asked Nao to repeat itself, and one lent in to see and hear the robot. While the robot asked the first questions, we observed participants laughing and smiling (6), two participants appeared confused, while one was unengaged. When the robot made an animal noise seven participants laughed, one appeared surprised and amused, while the previously unengaged participant smiled slightly. When Nao again began asking questions, we observed our previously unengaged participant appeared to again lose interest, four participants smiled, two appeared frustrated, one thoughtful, and one participant became annoyed with Nao appearing to misunderstand. When Nao thanked participants and said goodbye most participants (5) responded with "goodbye", one with "Ok" while the remaining participants did not respond or react.

Post Human-Robot Interaction Interviews have shown, that when asked about the experience, the participants were positive, describing it as "enjoyable", "fun", "interesting" and "surprising". The robot appeared to be more interactive than they had expected. The participants have nearly unanimously expressed astonishment with the level of animation shown by Nao. Compared to the lively movements the communication was found lacking. Participants described problems understanding the robot and complained about needing to repeat phrases, which caused disruptions in the interaction. Several participants (P2, P5, P6, P7) expressed disappointment with the scripted nature of the interaction.

Participants anthropomorphised the robot; Nao was seen as cute in appearance; with small stature, big eyes, and adorable voice. They often referred to the robot using the pronoun 'he', P1 expressed surprise having referred to Nao in such manner: "Sometimes I find it so quiet; I wonder what he… well, that's interesting, I think 'he', not 'it' …" [P1].

One participant mentioned the internal conflict they felt when thinking about the robot and how, for some people, this conflict could cause a barrier to interact with robots. **"**I could feel that instinct in my head (...) to treat it the way that I would treat a human…" [P3]. It was noted that the robot was obviously programmed to be friendly and that its' cuteness might reduce the negative media portrayals which surround robots that are too humanlike (P4). Participants P1, P2, P4 mentioned the concept of the uncanny valley, emphasising that the robot has not yet bridged that gap, which works to its benefit, allowing it to elicit more trust.

When asked if they could see this or similar technology in a social sector, responses focused on educational (P3, P5, P9) or home environments (P4, P7). P7 suggested that a robot could be used around the home/office to complete mundane tasks. It was noted that children might enjoy the simple interactions and a robot might encourage those who have difficulty engaging with traditional educational modes.

Overall, the participants found the interaction a positive experience and the robot itself an interesting, if novel, tool. P9 stating "(…) it does endear itself to you quite quickly cause it's quite pleasant and entertaining." While P6 concluded, "He's adorable and I want to take him home."

5. DISCUSSION & CONCLUSIONS

With the rise of social robot development, the user acceptance by the average consumer during their personal lives raises new challenges and increases the need for investigations such as those we present here. We provide one of the few studies of human-robot trust that involved human participants directly interacting with a social robot, whilst utilising a wide mix of quantitative and qualitative measures to capture the data. Our participants had not previously interacted with a robot and therefore we gained insight into their pre-, and post-human-robot interaction perceptions. The research revealed that although the participants' experience of robots came mostly from media, science-fiction movies, and dystopian literature, where they have damaging portrayals, it did not result in negative attitudes. Furthermore, although the participants pointed out that the child-like physical appearance of the robot was designed to be appealing and entice them, this did not reduce their levels of trust.

Lastly, the recurring concern among participants is not the robot itself, but the programmer behind the code.

Given the increased trust perception by participants presented by the TPS score and the increases in perceived Anthropomorphism and Animacy presented by the Godspeed scores, further investigation of the relationship of trust and physical appearance and movement is warranted for social robots. Interestingly before interacting with the robot the Perceived Intelligence and Safety scores on the TPS questionnaire received the highest scores. While these did not increase after the interaction, they remained positive suggesting that even given negative perceptions discussed based on pop-culture references, participants before and after interaction were generally accepting of social robots.

The interviews conducted after the interaction with Nao have shown all participants being surprised by the robot's lively, and animated persona. The initial reactions noted during observations were positive, expressing interest, amusement, and joy, with a hint of uncertainty. All of the participants commented on the unexpected amount of movement, and life-like reactions. However, all have also discussed dissonance between human-like behaviours and underwhelming conversational skills. Participants were frustrated not experiencing social cues, and needing to repeat themselves, some interpreted the interactions as ‘abrupt’. On the other hand, those negative impressions were mitigated by the overall cuteness, and childlike physical appearance of Nao, leading participants to anthropomorphize it, [17][18][19], referring to it as ‘he’, smiling at it, assuming intentionality in its behaviours.

Our mixed-methods approach provided a rich data set that shows promise for use with a larger sample in future work. The short interactions were enjoyable and delightful for participants, although, some traces of hesitancy and angst were detected. Longer and more complex interactions may provide both deeper understandings for the first-time users, illustrating capabilities of the social robots as well as further opportunities for investigating emotions. The rich data provided by video analysis of emotional cues appear unique to our study and, we trust, warrants further consideration. Lab setup to ensure successful capture of this, and other biometric data is required to provide new insights to the field of HRI. Finally, a more comprehensive pre- and post-interview may facilitate deeper analysis providing further insights into first-time human-robot interaction perceptions, vital in sustaining successful relationships with robots.

6. REFERENCES

[1] Lotz, V., Himmel, S., & Ziefle, M. (2019). You’re my mate–acceptance factors for human-robot collaboration in industry. In International Conference on Competitive Manufacturing, Stellenbosch, South Africa (p. 10).

[2] Kildal, J., Tellaeche, A., Fernández, I., & Maurtua, I. (2018). Potential users' key concerns and expectations for the adoption of cobots. Procedia CIRP, 72, 21-26.

[3] El Zaatari, S., Marei, M., Li, W., & Usman, Z. (2019). Cobot programming for collaborative industrial tasks: an overview. Robotics and Autonomous Systems, 116, 162-180.

[4] Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust 1. Journal of personality, 35(4), 651-665.

[5] Schaefer, K. E., Chen, J. Y., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. Human factors, 58(3), 377-400.

[6] Billings, D. R., Schaefer, K. E., Chen, J. Y., & Hancock, P. A. (2012, March). Human-robot interaction: developing trust in robots. In Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction (pp. 109-110).

[7] Sanders, T., Kaplan, A., Koch, R., Schwartz, M., & Hancock, P. A. (2019). The relationship between trust and use choice in human-robot interaction. Human factors, 61(4), 614-626.

[8] Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human factors, 57(3), 407-434.

[9] Marsh, S., & Dibben, M. R. (2003). The role of trust in information science and technology. Annual Review of Information Science and Technology, 37(1), 465- 498.

[10] Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. ACM Transactions on management information systems (TMIS), 2(2), 1-25.

[11] Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: the influence of emotion on trust. Journal of personality and social psychology, 88(5), 736.

[12] Tussyadiah, I. P., Zach, F. J., & Wang, J. (2020). Do travelers trust intelligent service robots?. Annals of Tourism Research, 81(C).

[13] Recio, D. L., Segura, L. M., Segura, E. M., & Waern, A. (2013, March). The NAO models for the elderly. In 2013 8th ACM/IEEE international conference on human-robot interaction (HRI) (pp. 187-188). IEEE.

[14] Andreasson, R., Alenljung, B., Billing, E., & Lowe, R. (2018). Affective touch in human-robot interaction: conveying emotion to the Nao robot. International Journal of Social Robotics, 10(4), 473-491.

[15] Rosi, A., Dall’Asta, M., Brighenti, F., Del Rio, D., Volta, E., Baroni, I., ... & Scazzina, F. (2016). The

use of new technologies for nutritional education in primary schools: a pilot study. Public Health, 140, 50-55.

[16] Ioannou, A., & Andreva, A. (2019, September). Play and Learn with an Intelligent Robot: Enhancing the Therapy of Hearing-Impaired Children. In IFIP Conference on Human-Computer Interaction (pp. 436-452). Springer, Cham.

[17] Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. Robotics and autonomous systems, 42(3-4), 143-166.

[18] Fink, J. (2012, October). Anthropomorphism and human likeness in the design of robots and human-robot interaction. In International Conference on Social Robotics (pp. 199-208). Springer, Berlin, Heidelberg.

[19] Goetz, J., Kiesler, S., & Powers, A. (2003, November). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. (pp. 55-60). IEEE.

[20] Schaefer, K. E. (2016). Measuring trust in human robot interactions: Development of the “trust perception scale-HRI”. In Robust Intelligence and Trust in Autonomous Systems (pp. 191-218). Springer, Boston, MA.

[21] Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International journal of social robotics, 1(1), 71-81.

[22] Sanders, T. L., MacArthur, K., Volante, W., Hancock, G., MacGillivray, T., Shugars, W., & Hancock, P. A. (2017, September). Trust and prior experience in human-robot interaction. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 61, No. 1, pp. 1809-1813). Sage CA: Los Angeles, CA: SAGE Publications

Teleworker's Perception of Technology Use for Collaborative and Social During the COVID-19 Pandemic

Jie Cai, Sarah J Ryu, Donghee Yvette Wohn, Hyejin Hannah Kum-Biocca
New Jersey Institute of Technology

New Jersey, United States

jie.cai@njit.edu; sr984@njit.edu; wohn@njit.edu, hyejin.kum-biocca@njit.edu

With the flourish of collaborative and social technologies in the market since the pandemic, there is limited understanding of user's attitudes towards these technologies. We aim to understand teleworkers' perceptions of technology use during the pandemic and interviewed 46 teleworkers. We found that teleworkers generally hold a positive attitude towards social technologies and are creative to use these technologies to meet their social needs; they express overall negative feelings about remote collaboration technologies, though online communication flattens the communication hierarchy in the organization. The pandemic amplifies the extant challenges and highlights the shortcomings of technological design in well-established teleworking research and remote collaboration work. We suggest that future design should 1) combine and commercialize solutions that are well-grounded in prior work; 2) consider scenarios that are typically missed and can be easily replaced with collocated interaction from the pre-pandemic context into the forced teleworking context.

COVID-19 pandemic, work from home, social and work technology, remote collaboration

INTRODUCTION

The global coronavirus 2019 (COVID-19) pandemic has led to government-mandated physical distancing and even lockdown in some areas. Many organizations scaled back their business or even ceased operations in the regular workplace, instead, asking employees to work from home (teleworking) to reduce the infection risk (Mattern et al. 2021). Additionally, simulations show that many areas in the world will probably suffer several waves of contagion and recurring lockdown (Kissler et al. 2020). The world-spread pandemic with the government-imposed restriction not only affect healthcare and economies at the collective level, but also change the lifestyles of people at the individual level, and these changes have become the new normal of daily life (Liu et al. 2020).

The pandemic accelerates the trend to migrate to teleworking with appropriate technology use in both social and work contexts and increases our dependency on technology use (Waizenegger et al. 2020). For example, the downloading of video conferencing software jumped 45% immediately after the early lockdown and reached a 90% average increase in downloads compared to pre-COVID demand (Sydow 2020); most people reported that their social media consumption and posting have increased during the pandemic (Wold 2021). Though a lot of news and reports show the increase of technology use during the pandemic and the role technology played in transforming businesses or improving the psychological treatment of patients online (Saladino et al. 2020; Nah and Siau 2020), revealing the promising and positive aspect of technology use in this new normal, there is limited knowledge about user's perception about the role of technology played in work collaboration and social interaction during the pandemic. We acknowledge the well-grounded research in telepresence and remote collaboration in Human-Computer Interaction (HCI) and Computer-Supported Corporate Work (CSCW) (e.g., Kuzuoka 1992; Ishii et al. 1993; Brave et al. 1998; Fussell et al. 2004). However, prior work has not been conducted in a pandemic context with forced quarantine and physical distancing. It is still unclear to what extent prior work can apply to this new normal context and what still needs to improve to adapt technology to this context. Thus, it is critical and timely to understand user's perceptions of the current technology use to help researchers

and designers identify the challenges of technology use in the enforced teleworking environment.

In line with recent work proposing the necessity to review the technology use during the pandemic and considering them continuously used over time or cyclically in the recurring outbreaks (Golinelli et al. 2020), in this paper, we interviewed 46 teleworkers working/socializing during the pandemic to explore how users collaboratively use technology in both social and work contexts and to explore their perceptions and attitudes towards the role that technology played. We show how teleworkers perceive the technology use, praising the positive sides while also highlighting the challenges during the usage. These findings highlight the existing challenges in commercialized collaboration tools in well-established research about teleworking and remote collaboration, support recent work proposing the use of new and positive technology to mitigate COVID-19 impact on individual users, and imply to customize and improve current technical solutions, fulfilling people's needs in different contexts during the pandemic and beyond.

RELATED WORK

The pandemic accelerates the digital transformation in many domains, such as healthcare and business (e.g., Golinelli et al. 2020; Nah and Siau 2020). In this section, we focus on teleworkers' experience and perception towards technology use in remote collaboration and socialization.

Teleworking and COVID-19

Prior work on teleworking has discussed a wide range of issues such as work-life boundary management (Fonner and Stache 2012), advantages (e.g., perceived autonomy, job satisfaction, performance, and role stress) and disadvantages (e.g., feelings of alienation, isolation, worry) of teleworking (Collins 2005; Gajendran and Harrison 2007), and technological impact on teleworking (Cox et al. 2014).

The role of technology in teleworking can be both positive and negative. For example, mobile technology allows people to work remotely with flexible hours that fit around caring responsibilities; it also results in an "always-online" culture that can disrupt work-life balance and induce stress regarding constant connectivity and responsiveness (Cox et al. 2014). Stawarz et al. (2013) found that though tablets can facilitate both home and work tasks, they can potentially blur the work-life boundary by encouraging people to do work tasks during home time and vice versa. Prior work also suggests that controlling working hours is the most critical ability to achieve a positive work-life balance (Maruyama et al. 2009), but teleworkers grapple with the tension between the desire for flexibility and the need to maintain the work-life boundary (Fonner and Stache 2012).

Olson and Olson (2000) explain that effective teleworking in the sociotechnical condition is determined by four key components: common ground, coupling/dependency of work, collaboration readiness, and collaboration technology readiness. They suggest that groups, with high common ground and loosely coupled work, with readiness both for collaboration and collaboration technology, have the potential to succeed with teleworking, and deviations from each component restrict teammates' performance and require changes during the collaboration.

Different from regular teleworking, the pandemic quarantines people, forces people to work from home, restricts mobility, and enforces physical dis- tancing; thus, the enforced teleworking causes chal- lenges to communication and personal well-being (Waizenegger et al. 2020). Forced teleworking might change the interaction dynamics of the four components by Olson and Olson (2000). For example, teleworkers need to adapt quickly to unfamiliar digital technologies to maintain regular business and communication. The enforced teleworking with high intensity can deteriorate the relationships with coworkers as well (Gajendran and Harrison 2007).

Remote Collaboration and Technology Use during COVID-19

Remote collaboration research is well-grounded in both HCI and CSCW. Though distance matters, lots of work has explored paradigms and prototypes to overcome the shortcomings of distance and physical presence in decades. For example, in the 1990s, Kuzuoka et al have developed prototypes about video-mediated communication systems to stimulate spatiality and deliver emotional expression (e.g., Kuzuoka 1992; Kuzuoka et al. 1994). A group of researchers has focused on tangible interfaces to overcome the haptic limitation and physical background in graphical user interfaces and to enhance remote collaboration and communication (e.g., Ishii and Ullmer 1997; Brave et al. 1998). Research about gaze motion and gesture and shape capture and their implications on telecommunication and telepresence has been continuously discussed (e.g., Ishii et al. 1993; Fussell et al. 2003; Sirkin et al. 2011; Leithinger et al. 2014).

With the advancement of technological infrastructure, remote collaboration research has been involved in the mobile collaboration (Gauglitz et al. 2012), information visualization (Balakrishnan et al.

2008), Augment Reality (AR) (Gauglitz et al.2014a), Virtual Reality (VR) (Williamson et al. 2021), and Mix Reality (MR) (Piumsomboon et al. 2018; Teo et al. 2019). For example, Gauglitz et al. (2014b) has developed a live AR-based prototype allowing a remote user to mark annotations through a touchscreen to instruct and assist a field worker to solve problems. All these novel techniques are still at their early stage in lab or field research and not available at scale in commercialized remote collaboration platforms.

The current crisis increases society's reliance on technology with social and occupational changes (Garfin 2020). Many media reports show a surge of video conferencing app downloads for work collaboration and keeping in touch with family and friends in Europe and the US (Sydow 2020; Trueman 2020). The transformation from workspace to teleworking decreases the frequency of ad- hoc meetings due to the lack of spontaneity and further impedes knowledge sharing and coordination for knowledge workers (Waizenegger et al. 2020). The increased video conferencing meetings can induce fatigue and anxiety towards the overuse of technology (Wiederhold 2020). The technological exhaustion results from the technological shortcomings of video conferencing platforms such as lack of eye contact, and limited nonverbal cues (Riva et al. 2020). Wu and Chen (2020) found that the technology use also increases workload and decreases productivity and that the technical issues, such as less efficiency of online communication, are the core reasons. Though some work points out the limitations of this enforced teleworking context, there still lacks a focused and comprehensive study to explore teleworkers' perceptions of technology use for work collaboration during the pandemic. Thus, we asked:

- **RQ1:** What are teleworkers' perceptions of using technology to work and collaborate with others remotely during the pandemic?

As for social context, research proposes to use technology to improve the quality of our personal experience and to augment the existing strategies to generate psychological well-being; social technology such as m-Health and smartphone apps, stand-alone and social virtual reality, video games, exergames, can potentially enhance our personal experience (e.g., emotion, engagement, and connectedness) that is challenged by the pandemic (Riva et al. 2020). For example, video games allow individuals to connect and socialize through play as a way to mitigate the psychological impact of the pandemic (Marston and Kowert 2020). Exergames are easy-to-use tools to cope with social isolation and maintain physical fitness during the quarantine (Viana and De Lira 2020). In general, video games not only provide ways to connect with others and maintain mental well-being, but also offer opportunities to substitute reality to host events and to find ways to gamify everyday activities (Kleinman et al. 2021). Research targeting aging populations' technology use found that older adults use various mobile technologies to stay connected with family (e.g., Facetime, Skype), maintain mobility, and link to resources helpful for their physical and psychological well-being (Banskota et al. 2020). While many media reports show the increased use of mobile technology and social media, and some research highlights the importance of different technologies that can potentially benefit the users, little user research explores teleworkers' attitudes and perceptions of the technology use in social context during the pandemic. Thus, we asked:

- **RQ2:** What are teleworkers' perceptions of using technology to socialize with others remotely during the pandemic?

METHODS

Participant Recruitment

We received approval from the Institutional Review Board (IRB) to carry out the project and interview protocol. We interviewed 46 participants who were adults from the United States who had to work from home due to the pandemic. The participants were recruited through a convenience sample and were family, friends, or loose acquaintances of our research assistants in the lab and 12 students taking a computing-related course. Most participants were young adults in their 20's (60.9%) (M = 27.7, range = 19 to 55). There were 19 female participants (41.3%) and 22 male participants (47.8%). There were 23 Asian participants (50.0%),10 Caucasian participants (21.7%), 5 African American participants (10.9%), and 3 Hispanic participants (6.5%). Five participants, labeled as N/A, did not share their demographics (10.9%). Demographic details and participant occupation details are summarized in theAppendix A.

Interview Process

Semi-structured interviews were conducted remotely via Skype, Discord, or FaceTime. They lasted for around 15 to 45 minutes. We started with fixed warm-up questions about their general remote living experience like *"What does your current living situation look like?"* and *"Can you walk me through what you did yesterday?"* Then, we asked questions more specific to our research questions (This part of the protocol is semi-structured with fixed questions

related to work and socialization and potential follow-up questions based on participants' responses). We first asked questions about their remote work transition with questions like *"Do you think the transition to work from home was easy or difficult?"* and *"Has this work environment affected your productivity?"*, following up with questions about their collaboration experience like *"Has the collaboration method changed since the quarantine"* and *"How is the new method of collaboration impacting your productivity?"* Next, we asked questions about their remote lifestyle transition with questions like *"How often do you go out of your home?"* and *"What events have been canceled for you?*", following up with questions about their socialization experiences like *"How do you connect with family and friends virtually?"* and *"What are the pros and cons in streaming holiday parties and events?"*. Throughout the interview, we asked participants to elaborate on their thoughts and feelings. Finally, we asked general demographic questions about their age, race, gender, and occupation. We also opened the floor to any other related insights or experiences the participants may have had that were not covered during the interview.

Interview Analysis

Each interview recording was transcribed through speech transcription software. The research assistants then reviewed and familiarized themselves with each transcript. We used a thematic analysis to identify, classify, and analyze themes (Nowell et al. 2017) shared within participant's remote in- teraction experiences during the pandemic. To find these themes, we used an inductive coding pro- cess. Five researchers were involved in the analysis process (three for coding and organizing and two for calibration). First, three researchers organized participant's responses on a spreadsheet, based on their relevance to the research questions. Sec- ond, each researcher developed codes from the organized responses, which were presented and explained in detail during weekly meetings with the other two researchers. During these meetings, the team reviewed and discussed the codes and quotes together until all five researchers reached an agreement. This process was repeated until all codes were consistent and finalized for both research questions. Third, after the coding process was completed, these codes were imported into Miro, an online collaborative whiteboard tool. Each code became a sticky note on the whiteboard, with a tag that displayed which participant it belonged to. Three researchers worked together to group codes and presented themes found among the grouped codes with the other two researchers for calibration. Fourth, when the major themes were identified, we reviewed each theme's codes in-depth to find any sub-themes. All the themes and sub-themes were developed and agreed upon by the five researchers. Finally, quotes that best represented the themes were found and reported in the results section. This analysis process took several weeks to complete through many weekly meetings and discussions.

RESULTS

Perceptions of Work Collaboration

Flatten Communication Hierarchy

As an outcome of transitioning to online communication, participants developed more streamlined communication channels with others. Before the pandemic, P33 (M, 26), a software engineer, would ask questions or receive updates indirectly through his co-worker, who relayed the information from their boss. However, as a result of being remote, P33 now communicates with his boss, which he noted was a significant change in his collaboration. Sim- ilarly, P38 (M, 29), a university lecturer, found that teaching classes online through WebEx, provided more opportunities for his students to ask him private questions without interrupting the class, which was challenging to do in person. The straightforward nature of communication platforms allowed participants to communicate more directly with others, especially those higher on the chain of command.

Extend Workload and Duration via Remote Collaboration

Tasks that were more individually focused were suited to be done remotely; however, tasks that require collaboration were more difficult to do remotely in comparison, extending workload and completion time as a result. Being in the office naturally facilitates conversations among co-workers to solve any spontaneous problems or provide updates. However, because participants working from home needed to wait for responses or scheduled meetings to collaborate, it added more effort and time to complete tasks or notify others. P34 (F, 30), a corporate office manager for a car wash, found it challenging to stay connected with her team. It took extra work to set up multi-way calls and to call her team for every update: *"If we were right next to each other, we could just say, 'Hey, this is what's going on today and that's it. But now it's a little different because things tend to pop up throughout the day where we need to call each other."* Remote collaboration does not work as smoothly as an in-person collaboration due to the delay in reaction from others and additional effort to stay in touch.

Blurs Time on and off-hours

Working at the office allowed participants to create boundaries between work and personal life

due to the physical distinction between the two locations. However, remote work solutions increased accessibility to work, blurring the lines between work and personal life for those who worked from home. As a result, participants were expected to work beyond their standard work hours, which interfered with their personal schedules: *"I rarely received information from the team members at seven or eight in the evening, but now the information that needs to be responded to during this period has increased"* (P30, M, 31). P17 (F, 55) was appreciated by her supervisor for putting more effort to be more accessible '24/7' to her students and parents during the pandemic.

More Tracking and Planning Needed

More planning and tracking were required to continue working from home. P46 (M, 23) is a network administrator, developer, and recruiter: *"Arranging [meetings] for more than two people, including myself, to collaborate requires a little more planning compared to before, where I could just walk over to the desk of a colleague and see if they were busy and just start talking to them."* Just to collaborate with others, extra planning and setup were needed, which slowed work efficiency. There were also more tools participants needed to keep track of in order to work and collaborate remotely.

Technical Tools Delayed Productivity

Many adopted new technologies to work from home; however, managing multiple platforms became more difficult with the lack of experience using them. P29 (F, 26), a geometry teacher, felt that her job became more strenuous as a result of all the new tools she had to 'keep up with' to continue teaching online. Along with the limited experience using these technologies, technical difficulties also slowed work efficiency: *"Sometimes it's faster when it is in-person because you are able to show the person additional information right in front of them. With online [tools], technical issues always seem to cause delays, such as getting the shared screen setup or someone talking while on mute"* (P41, F, 26). Because participants had to find technological solutions at such short notice, the lack of expertise and training to use these tools made it difficult to work remotely.

Lack of surveillance

Participants reported a decrease in surveillance from their supervisors and coworkers, which created less pressure while doing their work: *"There's no one right behind you that's looking at you constantly and making sure that you're doing work"* (P16, M, 22). Some participants reacted to the lack of supervision and decreased pressure positively. P16 (M, 22) initially expected the transition to work remotely would bring more strict oversight. He was surprised to experience the opposite effect. As a result, P16 felt it allowed him to work at his own pace due to the decrease in pressure.

Others reacted to the lack of supervision negatively, feeling that their work was not adequately acknowledged. P20 (M, 24), a reliability engineer, found it difficult to stand out and be recognized for his work when his supervisors were not present. Similarly, P9 (M, 31) expressed how he felt less appreciated and valued when his actions were not seen directly by his boss. While the lack of surveillance lightened the weight of expectations from supervisors, it also made it easier for many efforts to be overlooked and underappreciated. Some participants felt less motivated to stay on top of their work (P20, M, 24).

Lack of Physical Presence

Lack of Proximity

Participants missed being in the close presence of their co-workers. P1 (M, 21) is a receptionist at a dental office who was suddenly conscious of the lack of his co-workers' presence. He missed having them around, describing them as 'good' and 'fun'. Other participants similarly longed for face-to-face interactions with their co- workers and were unaccustomed to not having them around. While participants appreciate having the technology to continue meeting with their co- workers, online interactions could not substitute face- to-face interactions: *"I thank God [that] we have this internet world that we could depend on, but you know, I'd like to see them face to face"* (P4, M, 23). Working from home made distancing from co- workers inevitable.

Lack of Spontaneous Collaboration

Many participants preferred working in the office because being near other co-workers allowed for spontaneous collaboration. Through body language and casual conversation, participants could gauge when it was the right time to approach a coworker to facilitate an unscheduled meeting. These impromptu conversations in the office made it easier to get instant feedback and responses. It was difficult to replicate the same experience online because remote collaboration was very planned and structured, leaving little room for spontaneity. Remote participants needed to create meetings or wait for others to respond: *"I guess asking for help and stuff is a little difficult because you have to set up an appointment or set a meeting with my boss and like whoever can help me and make sure that they're free. And if they're not free, I have to wait for them. Meanwhile in the office, if I saw they were free I could just ask them. It's a lot more difficult to collaborate"* (P16, M, 22). Collaborating online slowed down people's ability to work out problems together and limited who could

get involved. P29 (F, 26), a geometry teacher, got more work done collaborating with other teachers in-person compared to sending emails back and forth: *"It's better and more efficient when teachers meet to collaborate on upcoming events because it takes less time and more people can participate."* Overall, the lack of a method to spontaneously and seamlessly collaborate with others hindered participant's ability to work as efficiently.

Lack of Idea Generation

It became more difficult to develop new ideas and be creative due to the inability to freely exchange ideas and thought processes with others. This was further hindered by the limited time of scheduled meetings and wait time for responses from others. P32 (F, 33), is an attorney: *"If it was purely online, I think it would be harder to figure out creative solutions to problems because you don't have the back and forth that a conversation would allow you to, uh, not have the bounds."* Because meetings are more planned, they are limited in allowing extra time to expand on ideas or conversations. P7 (F, 22), a process engineer, found that meetings became more straight to the point, making it difficult to develop ideas further or consider alternatives when the team was set on a particular solution.

Lack of Emotional Expression and Shared Understanding

Body language and emotional cues were difficult to communicate online, which hindered the participant's ability to monitor their coworker's reactions as feedback. P30 (M, 31) used his coworkers' expressions to guide decisions and check if his team was being productive: *"It's hard to feel everyone's emotions through remote video... I need to always pay attention to everyone's reactions, which will make it easier for me to make decisions."* Others, like P41 (F, 36), lost the ability to use reactions to measure whether the team needed more elaboration on material: *"Since we do not do video conference for meetings, you do lose the ability to see people reactions. By reading certain people's faces in meetings I am able to potentially expand on certain subject matters within the meetings."* Overall, online collaboration tools could not translate nonverbal communication smoothly, like body movements or facial expressions. It puts participants at a disadvantage because it was a viable source of information they could use to help them collaborate.

Lack of Clarity of Expression via Video

Video calls made it difficult to clearly express and convey information to others during meetings. For example, current remote work solutions were limited in functionality to support those, like P46 (M, 23), who visually express themselves or are more hands-on: *"I'm a really visual person and hands-on, so already going completely online kind of puts me at a disadvantage really. So when I'm explaining something to someone, it gets a little harder to communicate exactly what you need when you're doing virtually versus just writing or drawing it out."* Before the pandemic, P21's (F, 22) mentor at a technology co-op spoke directly to her and personally explained projects and tasks. This allowed P21 to understand the material at her own pace and guide the conversation to her needs. However, when meetings were online, she found it difficult to follow the conversation when the calls became more generalized and indirect.

Meetings are Structured instead of Free-form

As a result of participant's limited time with other coworkers, meetings had become more structured to ensure work needs were met. P17 (F,55), a high school science teacher, described how teacher meetings with administration transformed from being free-formed group work and discussion to prepared agendas and presentations. Having more free-form meetings provided a fresh experience for the same work: *"That's always changing. Like there are people to talk to. It's more spontaneous. Like I can go up to another co-worker and be like, hey, what's up? How are you doing? Sometimes they will have impromptu meetings so while the tasks are the same, the spontaneous gave something new"* (P16, M, 22). These face-to-face interactions were opportunities to socialize and build work relationships. It is challenging to facilitate similar casual interactions when an online interaction is very planned.

Perceptions of Socialization

Appreciating Seeing Faces and Talking Virtually But Still Missing Aspects of Physical Presence

In relation to the pandemic, participants appreciated having online socialization methods available that helped them stay connected with others without needing to leave their homes. P26 (F, 49) did virtual workout calls every day with her friends and found it to be a positive and motivating experience. She enjoyed being in her friends' company, despite being apart. Participants P43 (N/A, N/A) and P46 (M, 23) were more optimistic considering the lack of an alternative, with P43 in particular sharing that, 'having something is always better than nothing.' However, while many had positive remote social experiences, participants still felt that some aspects of in-person interaction cannot be fully replaced by technology. Online socialization temporarily satisfied people's need to socialize with others, but was not a long-term solution.

While participants could see and communicate with others online, some aspects of in-person interaction

were more challenging to convey through the screen, like physical presence and immersion. Many tried to make the most of the experience, however, it was difficult to enjoy virtual social events to the same extent as in-person events. There was a lack of physical presence when interacting with others digitally. P9 (M, 31) compared the presence of others with talking to plastic: *"And cons is, you feel physical plastic, you're not there with people, you can't celebrate with them directly... It's not the same."* Body language was also difficult to communicate online through a webcam, making it harder to measure whether someone was engaged in the conversation.

Socializing remotely does not feel completely immersive. Participant P1 (M, 21) enjoyed social events at the cost of not being fully immersed in the experience and environment. There was also a lack of expected preparation that comes with online socialization that contributes to its immersion. P2 (F, 22) found that not dressing up took the fun away from attending events. It was harder for participants to 'live in the moment' without having a completely immersive experience.

Accepting Limits of Main Social Media and Exploring Novel Ways

Despite the limitations of social technology, participants used it to their advantage to improvise new ways to spend time with family and friends.

Novel Apps to Socialize and Entertain
Participants used this time as an opportunity to explore and experiment with the diverse portfolio of socialization technology available to find unique ways to interact with friends and family. Simply having conversations on online communication apps was no longer the only option people had to socialize. Novel gaming apps, like JackBox, Houseparty, and GamePigeon, became more popular amongst participants to use during social hangouts with others because the gaming element added an extra layer of engagement to the experience. Many enjoyed having a more exciting method to interact and connect with others while social distancing.

Forgiving about Cancellations

Participants were more forgiving about event cancellations as a result of a shared duty to follow social distancing regulations and protect their own and others' safety. Many recognized that there will always be more opportunities to celebrate and attend events after the pandemic. P20 (M, 24), in particular, had a more positive outlook despite losing the experience of an in-person graduation ceremony for his daughter: *"I was even saving my sick days to be able to take off to go see her but it's okay, life is more important. We can celebrate these things later on."* Participants also adopted a cautionary mentality when approaching these cancellations, preferring to be 'safe than sorry' (P1, M, 21). Though participants were upset and disappointed, there were 'good reasons to cancel' (P7, F, 22) and it 'made sense'.

While there were virtual substitutes for these canceled events, participants were less inclined to attend because certain elements of the in-person experience could not be replicated. P15 (F, 24) felt less motivated to attend an online concert because it was not as full and immersive. However, participants were still accepting of these virtual alternatives because they were better than having nothing (P10, M, 46).

Exploring New Ways to Socialize

Participants utilized the available communication and socialization platforms to explore new ways they could host social events online. They celebrated virtual events like birthdays, happy hours with coworkers, and baby showers, among others. P32 (F, 33) and P27 (M, 35) both attended virtual funerals. P32 attended a Shivah that was streamed on Zoom to pay respects to her friend's grandmother. P27 felt that the virtual format for the Jewish funeral actually helped people reflect on the positive: *"It was really focused on saying good stories of her. And just then a bunch of people was laughing because they were funny and happy stories about her. So I think it was better than a normal funeral... People were happy because they were remembering good times."*

Some participants celebrated religious holidays virtually as well. P6 (M, 23) celebrated Ramadan through virtual streams held by mosques. P32 (F, 33) celebrated Seder with her family for Passover on Zoom. P13's (F, 23) family used FaceTime to help her celebrate the Tamil New Year tradition, despite living away from home: *"What they did was at 6:30 AM, they FaceTimed me. I had to somehow walk out of my room and into the sofa out here, eyes closed, literally, eyes closed, walked out, and somehow answer the FaceTime call, and I had to look at this and so that they flip the camera around so that, that's still the first thing that I saw would be myself. They set up the mirror and everything back in Edison and they did that and I still kind of appreciate it because they're still trying to do what normal traditions are."*

Some participants needed to get creative and improvise their technology use and their physical space for certain events. For her daughter's virtual ballet class, P42 (F, 33) used Zoom and accommodated her living space to make a makeshift ballet studio: *"For my daughter's ballet lesson, it was done through Zoom. We hooked it up to the TV in the living room so that she could have a big screen for it. I didn't want her to feel like she couldn't move too*

much and not be able to see the screen anymore. It worked out perfectly... Luckily we were able to figure out the TV situation for our daughter. We had never used it that way before. So it all worked out."

Relieves Some Concerns About Loved Ones Well-being

Participants were remotely communicating with family members and friends more often to check their well-being. While P15 (F, 24) lived alone before the pandemic, she had concerned family members contacting her often to see how she was doing. Others used these checkups to catch up and stay updated in each other's lives. These periodic well-being checkups helped relieve concerns about loved ones and were often the main motive for contacting them. However, there can be a technological divide between the younger and older generations.

Lower Barrier to Participate in Remote Socialization

Due to the quarantine and stay-at-home recommendations, there was a lower barrier for participants to join online social events with family and friends. More people could attend events because the online format saved commuting time, created fewer excuses to not attend, and was easier to arrange. Participants were also able to have more frequent and longer social events. P44 (F, 21) had longer back-to-back events with friends because they were all at home with nothing else to do. While she had organized online events before the pandemic, it would take longer to arrange due to delayed communication. Some participants felt that the stay-at-home situation allowed for more control over scheduling and for better availability. P44 (F, 21) was more motivated to reach out to friends because the quarantine provided a 'break' to do so. P45 (F, 21) similarly had more free time to check up on friends often, which improved her relationships with them. The increased accessibility of online communication and socialization tools, combined with people's increased availability brought by the pandemic, created more opportunities to socialize in a time where in-person activity was limited.

Facilitate Communication Among Those Who Don't Talk Regularly

Participants reported reconnecting with those whom they don't talk with regularly. P44 (F, 21) had tried to stay in touch with her longtime high school friends in the past but found it difficult to arrange a time within their busy schedules. The quarantine provided an opportunity for P44 to have a reunion with her friends as a result of the extra free time. P27 (M, 35) connected and networked with new people through a Zoom birthday party. The quarantine became a chance for many to rekindle old relationships or start new ones. Participants found themselves communicating more with others in general. By continually checking up on her family's well-being, P42 (F, 33) contacted her family much more often than before. P33 (M, 26) talked more with his parents and his extended family, like his aunts and uncles, who all live abroad. This does not only pertain to close family and friends; P45 (F, 21) found herself generally talking more with acquaintances through classes. Whether they have communicated often before the pandemic or not, participants found themselves initiating remote contact more often with family members, friends, or acquaintances in general.

DISCUSSION

This study could provide potential answers to recent research proposing to understand the adaption and impact of technology use during the pandemic by utilizing comparative analysis and to evaluate how teleworkers interact with technology and adapt to overcome contextual shortcomings (Dey et al. 2020).

By comparing the user's perceptions of technology use for work collaboration with their perceptions of technology use for personal socialization, we found that users generally had more positive experiences and attitudes for using social technologies. Users were understanding of the technological limitations and to some extent appreciated the availability of social technology, trying to find new online social platforms or re-purpose existing ones. However, users showed more negative attitudes and had less positive experiences with using collaborative technology. Users expressed strong concerns about the lack of physical presence and workload caused the technological limitations, though online communication flattens the communication hierarchy. Additionally, the overlap of concern between work and social technology use was the lack of physical presence, but it was a bigger concern in work collaboration than socialization. Possibly, task-oriented and relationship-oriented communication is the key difference. How the purpose of communication affects user's percep- tion of technology use might be an interesting topic to explore for future research.

The pandemic restrains user's mobility and affects their behaviors that normally have to do face-to-face. Users mixed the work and social technology use to meet their needs and overcome the shortcomings of not being able to gather together. For example, users reported using FaceTime to celebrate normal traditions with family remotely and attending funerals on Zoom, scenarios that have rarely happened online and neglected by designers in the pre-pandemic. The novel experience might reduce their

negative feelings and make them still consider these technologies helpful in general.

COVID-19 not only causes physical health concerns but also results in a number of mental health disorders such as stress, depression, and anxiety (Salari et al. 2020; Brooks et al. 2020). Psychological interventions related to providing social support are essential to preserving mental health, such as staying with family to relieve people's anxiety caused by the crisis (Cao et al. 2020; Son et al. 2020). Mental health professionals also suggest individuals use alternative communication methods such as social networks and digital communication platforms to cope with social isolation (Banerjee 2020), and propose to scale up mental health services and support through mobile technology interventions such as apps and text messaging (Figueroa and Aguilera 2020). We contextualize these suggestions and show that users appreciate the available social technologies to connect with family and friends, relieve their concerns, and interestingly, promote conversation among those who do not talk regularly. Is it a way to cope with anxiety, fear of death, or social isolation? What's the psychological logic behind this perception? It also raises questions about whether the crisis can impose communication and the potential consequences.

Teleworking in the pre-pandemic was described as "technologically feasible, flexible and autonomous, desirable and perhaps even inevitable, family- and community-friendly" (Bryant 2000). It seems that forced teleworking diminishes the benefits that are usually perceived in a regular/voluntary context. The overall negative attitude towards technology use in work collaboration can be in line with Olson and Olson's work (Olson and Olson 2000).

The pandemic makes people unable to be ready for collaboration. The new normal working styles need time to learn how to collaborate remotely, especially those who have not been working from home. The collaborative technology is still not ready to cope with the need to fully work from home as many users reported a lack of physical presence for the existing technologies. The homogeneous collaboration tools on the market still suffer many limitations that have been discussed and existed for decades (Ishii 1990; Kuzuoka et al. 1994; Ishii and Ullmer 1997). Though technology use increases, the dependency/ coupling of work is still strong. Typically, the strong dependency of work requires frequent and complex communication with short feedback loops and is usually achieved through collocated interaction (Olson and Olson 2000). The physical distancing and forced teleworking further amplify these challenges.

Common ground can be easily established in a collocated environment because people can share not only cultural and local context but also the more micro context of who is doing what at the moment and what remains to be done, and teleworking makes the establishment challenging because a variety of cues are lost (Olson and Olson 2000). We do not specifically address whether the users shared a common ground during the collaboration process. The issues (e.g., lack of clarity of expression, emotional expression, and understanding) might infer the difficulty to build common ground for teleworking collaboration, requiring further investigation.

Users showed mixed attitudes towards lack of surveillance, which, on one hand, reduced pressure from their supervisors; while, on the other hand, generated less appreciation from their supervisors. When the camera is not turned on, teleworkers' effort and progress were invisible to their peers and supervisors. These mixed attitudes support prior work stating that there is only a thin line of difference in perceiving the same technology for surveillance or for assessment (Das Swain et al. 2020). How to balance the support and surveillance of teleworkers should be evaluated further.

Design Implications

More Immersive and Embedded Environment of Collaborative Work

Many cues that are usually available in the workplace were missing from the forced teleworking. Though teleworkers can communicate and see faces virtually, the nuanced cues in physical environments are missed (e.g., social banter and buzz that constitutes an office environment (Mann and Holdsworth 2003)). A recent report also shows that being face-to-face is important not only for "getting the job done" but also for enjoying work more generally (Rubin et al. 2020).

The critical key to timely completing tasks is the fluidity of participation that allows one or two team members to work alone while others have a spontaneous meeting, team members moving from one subgroup to another or the team having a meeting as a whole (Olson and Olson 2000). The live synchronization of VR technology with more interactive tools/elements such as spatial audio, haptic feedback, 360 video streaming, and motion capture can potentially overcome the limitations of current communication systems in work collaboration (Neurons 2017; Huang et al. 2020). For example, is it possible to design a live VR meeting system with spatiality to simulate a virtual distance and facilitate the fluidity of online participation? The users can choose who to virtually 'sit' close with

and can have proximal interaction to stimulate the distance in the workplace. While someone is talking to close colleagues, it won't disrupt other colleague's performance, mimicking the social buzz in the physical environment. VR can also provide more cues to facilitate the common ground construction, reducing misinterpretations and the following work to repair. The Limitation of the head-mounted-display equipment in VR might also suggest that AR or MR technologies can have much more opportunities for innovation. We also want to clarify that these technologies were discussed early in research but have not been commercialized in a long time. The pandemic shows the necessity to consider feasible ways to make them available at scale.

Tools to Facilitate Peer Support and Self-evaluation and to Track Work Progress and Performance

Teleworking limits interactions with co-workers, and also makes it difficult to self-evaluate their performance and be aware of other co-workers' expectations during team collaboration (Das Swain et al. 2020). Nicholas Bloom (2020) suggests that an effective "performance review system" to evaluate teleworkers' output is essential to make teleworking succeed. Wu and Chen (2020) propose that HCI work should support on-site work and lab-based tasks to improve the feasibility of teleworking. In line with the thread, designers can design a collaborative tool to monitor user's work progress (not only the output) by checking in milestones and facilitate peer support so that co-workers can coordinate their tasks and normalize their routines to others', the supervisor can evaluate teleworkers' performances by not only the output but also the progress.

More Situated Elements for Socializing

Because users were more creative with social technology use and social activities were more fluid, there are opportunities to design tools that better support different social contexts. We exemplify that the users applied various technologies to socialize with others and re-purposed them to fulfill their social needs, situations that were not envisaged by the designers pre-pandemic (Dey et al. 2020). Future designs should consider how to stimulate and facilitate pre-pandemic physical group activities online. For example, is there a better way to support those attending funerals with family and friends with more engaging and emotionally connected tools?

Limitations and Future Work

This is a preliminary study about teleworkers' perceptions of technology use for work collaboration and socialization during the pandemic. It suffers several limitations. First, we used a convenience sample, so our findings may not be representative of all teleworkers. Future research should apply other sampling methods to increase the representation of the sampling. Additionally, the demographics of our samples show diverse professions. The pandemic affects different occupations differently. Future work can focus on a specific group and explore their perceptions in-depth. Second, our participants were individuals from the US and do not represent those from other countries, who may have faced different impacts and policies caused by the pandemic. The technological tools used for work collaboration and socialization may differ in other countries as well. Future work could involve a cross- country study to explore teleworkers' practices and perceptions across different cultural backgrounds and technological availability. Third, a majority of our participants were young adults, those who are older or not as technologically savvy were not well- represented. Future work could extend this study to explore older adults' perceptions toward technology use during the pandemic. Fourth, teleworking is also highly relevant to the family context, which was not focused on in this exploratory study. Future work can also consider exploring how the environment and family conditions affect their perceptions of socializing and working. Lastly, the data collection was completed in the early stage of the pandemic. It is an unusual situation for the participants to use technology for teleworking, where they might have experienced difficulty in using technology. Nowadays, they may have adapted to the technology use and changed their behaviors and perceptions towards the technology. Future work can also explore the dynamics of the perception of technology use during the pandemic.

CONCLUSION

This study aims to understand teleworkers' perceptions of technology use in work collaboration and socialization. We found that the challenges in work collaboration still exist and are amplified by the pandemic, though prior research has raised these questions and provided potential solutions. Users are creative to explore new technologies or re-purpose available technologies to meet social needs. Designers for telecommunication platforms should consider the implications of prototypes in prior work and find feasible ways to commercialize them, and designers should take the scenarios that are typically missed in social-technical design in the pre-pandemic into consideration.

ACKNOWLEDGMENTS

Thanks to Daniel for the assistance of data analysis and the research assistants from SocialXLab for data collection.

REFERENCES

Balakrishnan, A. D., S. R. Fussell, and S. Kiesler (2008). Do visualizations improve synchronous remote collaboration? In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, New York, New York, USA, pp. 1227. ACM Press.

Banerjee, D. (2020, 4). The COVID-19 outbreak: Crucial role the psychiatrists can play.

Banskota, S., M. Healy, and E. M. Goldberg (2020). 15 smartphone apps for older adults to use while in isolation during the Covid-19 pandemic.

Brave, S., W. Khii, and M. Hdre (1998). TangibleInterfaces for Remote Collaborationand Communication. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pp. 10.

Brooks, S. K., R. K. Webster, L. E. Smith, L. Woodland, S. Wessely, N. Greenberg, and G. J. Rubin (2020). The psychological impact of quarantine and how to reduce it: rapid review of the evidence.

Bryant, S. (2000, 3). At home on the electronic frontier: work, gender and the information highway. *New Technology, Work and Employment 15*(1), 19–33.

Cao, W., Z. Fang, G. Hou, M. Han, X. Xu, J. Dong, and J. Zheng (2020). The psychological impact of the COVID-19 epidemic on college students in China. *Psychiatry Research 287*.

Collins, M. (2005, 7). The (not so simple) case for teleworking: a study at Lloyd's of London. *New Technology, Work and Employment 20*(2), 115–132.

Cox, A. L., S. Dray, J. Bird, A. Peters, N. Mauthner, and E. Collins (2014). Socio-technical practices and work-home boundaries. In *MobileHCI 2014 - Proceedings of the 16th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 581–584.

Das Swain, V., K. Saha, G. D. Abowd, and M. De Choudhury (2020). Social media and ubiquitous technologies for remote worker wellbeing and productivity in a post-pandemic world. In *Proceedings - 2020 IEEE 2nd International Conference on Cognitive Machine Intelligence, CogMI 2020*, pp. 121–130.

Dey, B. L., W. Al-Karaghouli, and S. S. Muhammad (2020). Adoption, Adaptation, Use and Impact of Information Systems during Pandemic Time and Beyond: Research and Managerial Implications. *Information Systems Management 37* (4), 298–302.

Figueroa, C. A. and A. Aguilera (2020, 6). The Need for a Mental Health Technology Revolution in the COVID-19 Pandemic. *Frontiers in Psychiatry 11*.

Fonner, K. L. and L. C. Stache (2012). All in a day's work, at home: Teleworkers' management of micro role transitions and the work-home boundary. *New Technology, Work and Employment 27* (3), 242–257.

Fussell, S. R., L. D. Setlock, and R. E. Kraut (2003). Effects of head-mounted and scene- oriented video systems on remote collaboration on physical tasks. In *Proceedings of the conference on Human factors in computing systems - CHI '03*, New York, New York, USA, pp. 513. ACM Press.

Fussell, S. R., L. D. Setlock, J. Yang, J. Ou, E. Mauer, and A. D. Kramer (2004). Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction 19*(3), 273–309.

Gajendran, R. S. and D. A. Harrison (2007). The Good, the Bad, and the Unknown About Telecommuting: Meta-Analysis of Psychological Mediators and Individual Consequences. *Journal of Applied Psychology 92*(6), 1524–1541.

Garfin, D. R. (2020, 10). Technology as a coping tool during the coronavirus disease 2019 (COVID-19) pandemic: Implications and recommendations. *Stress and Health 36*(4), 555–559.

Gauglitz, S., C. Lee, M. Turk, and T. Höllerer (2012). Integrating the physical environment into mobile remote collaboration. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services - MobileHCI '12*, New York, New York, USA, pp. 241. ACM Press.

Gauglitz, S., B. Nuernberger, M. Turk, and T. Höllerer (2014a). In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*, pp. 197–205.

Gauglitz, S., B. Nuernberger, M. Turk, and T. Höllerer (2014b). World-stabilized annotations and virtual scene navigation for remote collaboration. *UIST 2014 - Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* , 449–460.

Golinelli, D., E. Boetto, G. Carullo, A. G. Nuzzolese, M. P. Landini, and M. P. Fantini (2020, 11). Adoption of Digital Technologies in Health Care During the COVID-19 Pandemic: Systematic Review of Early Scientific Literature. *Journal of Medical Internet Research 22*(11), 23.

Huang, H. Y., C. W. Ning, P. Y. Wang, J. H. Cheng, and L. P. Cheng (2020). Haptic-go-round: A Surrounding Platform for Encounter-type Haptics in Virtual Reality Experiences. In *Conference on Human Factors in Computing Systems - Proceedings*, pp. 10.

Ishii, H. (1990). Team workstation: Towards a seamless shared workspace. In *Proceedings of the 1990 ACM Conference on Computer-Supported Cooperative Work, CSCW 1990*, New York, New York, USA, pp. 13–26. ACM Press.

Ishii, H., M. Kobayashi, and J. Grudin (1993). Integration of Interpersonal Space and Shared Workspace: ClearBoard Design and Experiments. *ACM Transactions on Information Systems (TOIS) 11*(4), 349–375.

Ishii, H. and B. Ullmer (1997). Tangible bits: Towards seamless interfaces between people, bits and atoms. In *Conference on Human Factors in Computing Systems - Proceedings*, pp. 234–241.

Kissler, S. M., C. Tedijanto, E. Goldstein, Y. H. Grad, and M. Lipsitch (2020, 5). Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science 368*(6493), 860–868.

Kleinman, E., S. Chojnacki, and M. S. El-Nasr (2021). The Gang's All Here: How People Used Games to cope with COVID19 Quarantine. In *Proceedings of 2021 CHI Conference on Human Factors in Computing Systems - CHI ' 2021*.

Kuzuoka, H. (1992). Spatial workspace collaboration: a sharedview video support system for remote collaboration capability. In *Conference on Human Factors in Computing Systems - Proceedings*, pp. 533–540.

Kuzuoka, H., T. Kosuge, and M. Tanaka (1994). GestureCam: A video communication system for sympathetic remote collaboration. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, CSCW 1994*, New York, New York, USA, pp. 35–43. ACM Press.

Leithinger, D., S. Follmer, A. Olwal, and H. Ishii (2014, 10). Physical Telepresence: Shape capture and display for embodied, computer-mediated remote collaboration. In *UIST 2014 - Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, pp. 461–470. Association for Computing Machinery, Inc.

Liu, J., W.-H. Cheng, K. Nahrstedt, R. Jain, E. Ricci, and H. Byun (2020). Coping with Pandemics: Opportunities and Challenges for AI Multimedia in the "New Normal". In *ACM International Conference on Multimedia (MM '20)*, pp. 4794–4795. ACM.

Mann, S. and L. Holdsworth (2003, 11). The psychological impact of teleworking: stress, emotions and health. *New Technology, Work and Employment 18*(3), 196–211.

Marston, H. R. and R. Kowert (2020, 6). What role can videogames play in the COVID-19 pandemic? *Emerald Open Research 2*, 34.

Maruyama, T., P. G. Hopkinson, and P. W. James (2009). A multivariate analysis of work-life balance outcomes from a large-scale telework programme. *New Technology, Work and Employment 24*(1), 76–88.

Mattern, J., S. Lansmann, and J. A. Hüllmann (2021). It's not that bad! Perceived Stress of Knowledge Workers During Enforced Working From Home due to COVID-19. In *Proceedings of the 16th International Conference on Wirtschaftsinformatik (WI)*, Number March.

Nah, F. F. H. and K. Siau (2020). Covid-19 pandemic – role of technology in transforming business to the new normal. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 12427 LNCS*, 585–600.

Neurons (2017). How virtual reality facilitates social connection.

Nicholas Bloom (2020). How working from home works out. *Stanford, Institute for Economic Policy Research (SIEPR)* (June), 8.

Nowell, L. S., J. M. Norris, D. E. White, and N. J. Moules (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods 16*(1).

Olson, G. M. and J. S. Olson (2000). Distance Matters. *Human –Computer Interaction 15*(2-3), 139–178.

Piumsomboon, T., G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst (2018). Mini-me: An adaptive avatar for Mixed Reality remote collaboration. *Conference on Human Factors in Computing Systems - Proceedings 2018-April*.

Riva, G., F. Mantovani, and B. K. Wiederhold (2020, 9). Positive technology and COVID-19. *Cyberpsychology, Behavior, and Social Networking 23*(9), 581–587.

Rubin, O., A. Nikolaeva, S. Nello-Deakin, and M. t. Brommelstroet (2020). What can we learn from the COVID-19 pandemic about how people experience working from home and commuting? Technical report, University of Amsterdam.

Saladino, V., D. Algeri, and V. Auriemma (2020). The Psychological and Social Impact of Covid-19: New Perspectives of Well-Being. *Frontiers in Psychology 11*, 2550.

Salari, N., A. Hosseinian-Far, R. Jalali, A. Vaisi-Raygani, S. Rasoulpoor, M. Mohammadi, S. Rasoulpoor, and B. Khaledi-Paveh (2020). Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: A systematic review and meta-analysis. *Globalization and Health 16*(50), 11.

Sirkin, D., G. Venolia, J. Tang, G. Robertson, T. Kim, K. Inkpen, M. Sedlins, B. Lee, and M. Sinclair (2011). Motion and attention in a kinetic videoconferencing proxy. In *Lecture Notes in Computer Science*, pp. 162–180. Springer, Berlin, Heidelberg.

Son, C., S. Hegde, A. Smith, X. Wang, and F. Sasangohar (2020). Effects of COVID-19 on college students' mental health in the United States: Interview survey study. *Journal of Medical Internet Research 22*(9).

Stawarz, K., A. L. Cox, J. Bird, and R. Benedyk (2013). "I'd Sit at Home and Do Work Emails": How Tablets Affect the Work-Life Balance of Office Workers. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 1383–1388.

Sydow, L. (2020). Video Conferencing Apps Surge from Coronavirus Impact.

Teo, T., L. Lawrence, G. A. Lee, M. Billinghurst, and M. Adcock (2019). Mixed reality remote collaboration combining 360 video and 3D reconstruction. In *Conference on Human Factors in Computing Systems - Proceedings*, pp. 14. ACM.

Trueman, C. (2020). Pandemic leads to surge in video conferencing app downloads.

Viana, R. B. and C. A. B. De Lira (2020). Exergames as Coping Strategies for Anxiety Disorders during the COVID-19 Quarantine Period. *Games for Health Journal 9*(3), 147–149.

Waizenegger, L., B. McKenna, W. Cai, and T. Bendz (2020, 7). An affordance perspective of team collaboration and enforced working from home during COVID-19. *European Journal of Information Systems 29*(4), 429–442.

Wiederhold, B. K. (2020, 7). Connecting through Technology during the Coronavirus Disease 2019 Pandemic: Avoiding "zoom Fatigue". *Cyberpsychology, Behavior, and Social Networking 23*(7), 437–438.

Williamson, J. R., J. Li, V. Vinayagamoorthy, D. A. Shamma, and P. Cesar (2021). Proxemics and Social Interactions in an Instrumented Virtual Reality Workshop. In *Proceedings of 2021 CHI Conference on Human Factors in Computing Systems - CHI ' 2021*.

Wold, S. (2021, 9). COVID-19 is changing how, why and how much we're using social media. Technical report.

Wu, H. and Y. Chen (2020, 7). The impact of work from home (wfh) on workload and productivity in terms of different tasks and occupations. In *HCI International 2020 – Late Breaking Papers: Interaction, Knowledge and Social Media. HCII 2020. Lecture Notes in Computer Science*, Volume 12427, pp. 693–706. Springer, Cham.

APPENDIX: DEMOGRAPHIC AND INDUSTRY OF PARTICIPANTS

ID	Occupation & Industry	Age	Race	Gender
P1	Receptionist at a Dental Office	21	White	Male
P2	Customer Services Representative	22	White	Female
P3	Scheduling Manager	19	Asian	Male
P4	E-trader	23	White	Male
P5	Medical Advisor	N/A	N/A	N/A
P6	Assistant Project Manager	23	White	Male
P7	Process Engineer	22	Asian	Female
P8	Emergency Medical Technician	19	Asian	Male
P9	Desk Consultant and Fitness Employee	31	Hispanic	Male
P10	Library Staff	46	African American	Male
P11	Graphic Design Intern	21	Asian	Female
P12	Cybersecurity Technology At a Bank	51	Asian	Male
P13	Pediatric Newborn Research Assistant	23	Asian	Female
P14	Programmer & Business Owner	29	Asian	Male
P15	Production Engineering Intern	24	Asian	Female
P16	Data Analyst	22	Asian	Male
P17	High School Science Teacher	55	Asian	Female
P18	Software Developer	25	White	Male
P19	Coding Teacher	21	Asian	Male
P20	Reliability Engineer	24	Asian	Male
P21	Technology Co-op	22	Asian	Female
P22	Student	23	Asian	Female
P23	Unemployed	29	White	Male
P24	Senior Research Analyst in Healthcare	26	African American	Male
P25	RF Engineer for Corporate	27	Asian	Male
P26	Insurance Agent	49	Hispanic	Female
P27	Small Startup	35	White	Male
P28	Talent Coordinator	23	African American	Female
P29	Geometry Teacher	26	African American	Female
P30	Software Engineer	31	Asian	Male
P31	Software Engineer	26	Asian	Female
P32	Attorney	33	White	Female
P33	Software Engineer	26	Asian	Male
P34	Corporate Office Manager for Car Wash	30	White	Female
P35	Student	N/A	N/A	N/A
P36	Product Analyst	N/A	N/A	N/A
P37	Student	22	Asian	Female
P38	University Lecturer	29	White	Male
P39	Ph.D. student	25	Asian	Male
P40	Product Owner in Customer Technology Team	N/A	N/A	N/A
P41	Manager in Risk Adjustment for Healthcare Company	36	African American	Female
P42	Senior Tax Reporting and Company Analyst	33	Hispanic	Female
P43	Masters Student	N/A	N/A	N/A
P44	Student & Dance Teacher	21	Asian	Female
P45	Student & Substitute	21	Asian	Female
P46	Network Administrator, Developer, Recruiter	23	Asian	Male

User Engagement and Collaboration in the Next Generation Academic Libraries

Parisa Saadati
University of West London
London W5 5RF
parisa.saadati@uwl.ac.uk

Elena Pricope
University of West London
London W5 5RF
21438258@student.uwl.ac.uk

Jose Abdelnour-Nocera
University of West London
London W5 5RF
Jose.Abdelnour-Nocera@uwl.ac.uk

The dynamic progress in computing technology is driving the change towards a more centralised and connected society. The technology revolution has transformed cities into smart cities together with all the components. Various studies have shown an interdependence between the quality of life of citizens of smart cities and their level of education. It has been proved that Librarie's access can improve citizens' quality of life and has an essential part in people's education. Traditional libraries improve their services and adapt to the users' expectations using a push strategy, which focuses on predicting user expectations by improving library operations and information system functionality. However, the prediction is not always in line with reality when users are exposed to the most recent technology or forced to change by external forces such as COVID 19 outbreak. Intensive research on smart libraries and how automation can be implemented to support the activities of libraries is available from different backgrounds. The proposed research investigates how the next generation of academic libraries would improve user collaboration using a pull strategy method based on design thinking and a user-centered approach. The proposed research consists of field investigation on users' view in the University of West London library regarding user collaboration and using the survey findings to propose a solution. This paper includes a brief literature review on this topic with findings from previous studies and their limitations and proposed a methodology for collecting and analyzing the data.

Future scenarios, automated libraries, post-COVID transformation, user engagement.

1. INTRODUCTION

In the current global movement towards urbanisation, cities are now transforming into Smart Cities (SC). The term "smart" usually identified with the level of automation implemented in public and private sectors of a city and how well is Information Technology (IT) integrated into its activities. The main characteristics of SC are smart economy, smart living, smart governance, smart people, smart environment, and smart mobility (Ghosh and Mahesh, 2015). Available SC human aspects research shows a correlation between the education of citizens and the quality of life, and the overall quality of the city. (Ortiz-Fournier et al., 2010; Neirotti et al., 2014; Schopfel, 2018). Libraries are known to have an important role in education as they offer access to information and the physical space where people can interact and exchange ideas.

Libraries have changed over time to adapt to the changing needs of their users, such as ease the borrowing process and improve resource management. It is a common practice that library Information Systems (IS) are built on top of the legacy systems, integrating and upgrading on specific areas according to the requirements. Often, the upgrades are aimed to support the library staff in order to increase productivity and efficiency. This shows an inside-out approach aimed at the internal processes, and the library IS. The pace with which new solutions are emerging and the fact that library users are experiencing exposure to technology in other environments has become challenging to predict users' exact needs without their involvement. Furthermore, due to the lag between the idea generation and implementation, the new functionality is deployed by the time the users' preferences changed. Lack of communication between software developers and the users also influences this process.

The high amount of easily accessible online information reshaped the use of libraries in the past years from mainly providing access to information to providing a collaborative space for innovation by organising: social gatherings, networking events, targeted learning or entrepreneurship workshops (Leorke, Wyatt and McQuire, 2018). This shows an increased need in libraries IS to include functionality that allows users to collaborate, interact and communicate.

Covid-19 pandemic had a significant impact on libraries activities, especially on user collaboration and engagement due to the social distancing measures and other restrictions. Many libraries offered more online services or a hybrid approach where possible. As the situation stands, user collaboration in libraries depends on various IT solutions (available off-shelf) independent from the core library IS. This research is to propose a solution that can increase user collaboration in next-generation academic libraries. The following section looks at the current research studies in this area and their limitations.

2. LITERATURE REVIEW

The global value of the library automation market is estimated to reach $2.8 billion by 2024 (Hexa Research, 2017). Introducing RFID (Radio Frequency Identification Device) technology integration into libraries in various studies shows that it can automate storage, security, inventory, logistics, and delivery. Therefore, the users can use the library with minimal contact with the library staff (Snehalatha, Punjari and Gadgay, 2018). This novel technology uses radiofrequency tags and sensors to map the location of items, for example. In connection with the library system, the tag can have different values, allowing the user to borrow a book. The efficiency rate of RFID technology decreases when the tags and the sensors are located in the compacted areas. As most of the library's layout tends to be crowded and compact, this considers as one of the main limitations. In addition, to adopt RFID, a library must make a substantial investment to purchase all the equipment needed and integrated it into its system (Ozeer, Sungkur and Nagowah, 2019).

Onnasch et al. study investigated the impact of the Degree of Automation (DOA) on human performance. The research shows that when the DOA is high, the staff tend to be out of the loop due to reduced monitoring, over-trusting the software, and degrading skills. As a result, when the system fails, the impact is highly likely to be negative (2014). However, Tatasciore et al. study shows that high levels of DOA improve the operators' performance by reducing the workload and not affecting the level of situation awareness. The impact of DOA on operators (users) depends on how complex the tasks are automated (2020). These studies have been made in different environments; therefore, the findings reflect the applicability of automation in those situations.

In library management, automation can be the solution to many problems. However, implementing an automated system and its relevant impacts is a change for the libraries, which the users (staff and students) must support to succeed. Otherwise, it is most likely that this functionality will never use, and the project fails. Pre-implementation surveys proved a positive impact on user satisfaction as the user has direct input in the decision-making of the academic library improvements (Chow, 2019).

The development of strategies in libraries traditionally was based on push strategy where the focus is on improving the business activities' efficiency and predicting the users' needs from an in-out point of view. In recent years, due to the fast development of computing technology and its influence on users, this approach has been slowly replaced by the pull strategy, which has at its base innovative solutions centered towards the user and uses rapid incremental developments (Nelson, 2016). This shift allows the users to be more involved in the library developments, and it shortens the decision-making time, making the library more agile.

Design Thinking (DT) has improved the quality of applications with user-centric focused solutions in recent years. DT gives a systematic approach to the requirements specifications and increases the innovativeness for an existing approach using the end-user perspective. Therefore, the users' needs are not hypothesized, and instead, an on-field study will help determine what the user needs, how technology can help, and what is a more cost-effective solution (De and Vijayakumaran, 2019).

2.1 Users Involvement and Crowdsourcing

Library users play an essential part in academic libraries. Users' feedback typically improves the library's digital strategy, emphasizes user involvement, and aligns the libraries' operations with the users' expectations. The increased digitalisation in libraries has permitted the user and library to co-create value, so-called the value in use (e.g., feedback, applying filters). Islam et al. suggest that the service development process in libraries should focus on value co-creation, involving the students and library staff using interviews and surveys (2015).

Various studies support the implementation of crowdsourcing in libraries. The term was first introduced in the academic literature in 2006 by Howe (2006). He defined crowdsourcing as the act of assigning a task (performed by an employee) to a group of people as an open call. The practice involves the user in solving problems or contributing towards the development of various functionalities. Crowdsourcing uses the "mine the users" as a resource to add value to the knowledge generated in the digital library, improve the library services by involving the users in the process, and dimmish the timeframe in which goals are reached (Chhatwal and Mahajan, 2015). However, crowdsourcing adoption in academic libraries assumes mutual trust on both sides: the users and the library.

Trust is an essential factor in the smooth operation of the libraries. From a general perspective, the users' trust in the library staff will be reflected in the number of users, user interaction, and behavior. The trust level between staff members reflects in the daily activities. A high level of trust is associated with a low rate of sanctions, improves cooperation, increases engagement, enhances accountability, and a higher innovation rate (Wojciechowska, 2021). On the other hand, the disadvantages of using crowdsourcing are the complexity of managing a large group, users' lack of responsibility, and the quality of the solution generated by the crowd, which can be of low quality.

Crowdsourcing adoption in academic libraries can be achieved in a more controlled environment by imposing a private community of users. Therefore, users will be accountable more for their contributions and, at the same time, improve user collaboration in the digital space. Furthermore, using a reward system can motivate and engage the crowd.

User collaboration in libraries is not limited to student-to-student collaboration and group activities. It also covers collaboration between the library staff and the students, for example, in a face-to-face session where students can ask questions from staff and receive guidance on approaching a research topic or using most of the resources available.

The Butler and Byrd study on librarians' and students' perception shows that the students' perception of face-to-face research consultations with the academic librarian is rated higher than the self-evaluation of the librarians. Furthermore, the study shows that students' perception of the consultation is influenced more by the librarian-student interaction than by the accuracy of the provided information. On the other side, the librarians' perception is affected by the provider's pessimism as the librarians tend to underestimate the value of the consultation (2016).

Despite the current research on smart libraries and automation of library processes, there has been limited research on how IT can increase user collaboration in the next generation academic libraries in both digital and physical space.

3. PROBLEM STATEMENT

Covid-19 pandemic forces libraries to speed up adopting a hybrid operation model (dual operations on the digital and physical level). Academic libraries have moved into digital space in order to continue offering their services to users. This has affected the collaboration between students and the library staff due to government restrictions and social distancing regulations. In addition, user behaviour is influenced by new technological developments, as they are exposed to a new technology outside the library.

The proposed study is designed to recommend a framework or guideline that will increase user collaboration in the next generation of academic libraries based on survey findings and intensive literature review on current studies.

3.1 Research Question

This study intends to ask how next-generation academic libraries can increase user collaboration in their digital and physical space?

3.1.1. Research sub-questions

We also asked the three sub-questions below;

- What are the factors influencing user collaboration in academic libraries?
- What can be done to improve user engagement and collaboration in next-generation academic libraries?
- What level of automation will be appropriate to benefit user collaboration?

4. RESEARCH METHODOLOGY

4.1Overall Aim

The proposed research aims to recommend an IS framework to improve user collaboration in next-generation academic libraries (in digital and physical space). We believe that a list of factors will influence the relationship between users and the digital services in the libraries.

4.1.1. Research Objectives

- To investigate the UWL library users' perception of user collaboration.
- To analyse and evaluate the results of the questionnaire in order to find out what are the factors influencing the user collaboration.
- To make recommendations based on the findings.

This research will contribute towards the knowledge of academic libraries development using a pull strategy. Therefore, it can influence the academic library decision-makers, improve operational activities, and drive the change towards the next generation academic library.

4.2 Research Methodology

We will collect both secondary and primary data for this research; an extensive literature review to present the structured findings on the current studies on the topic and a survey evaluation on UWL students. In the library context, the student subjects from the university can be viewed as the major target of library services. Most participants in this study have prior experience of somehow automated services in UWL library. UWL library is not a fully automated library but offers some of the services of a smart library, such as a self-service kiosk, automated returns machine, and use of RFID tags.

The structured questionnaire will analyse and evaluate the UWL students' views regarding user

collaboration in the library. The questionnaire will be made available online for a better result. Hence participants can choose the most convenient time to complete the survey. In addition, this option will provide flexibility, and lockdown restrictions will not affect the research stage of the proposed project. This research will not collect or use participants' information or data, making the participants identifiable.

In order to appropriately measure the hypothesis, research items will draw and modify from well-validated sales. The advantages of using existing scales are that they fulfill the face and content validity, thus providing a theoretical basis for the assessment domain.

Once the empirical data is collected, the statistical analysis will be made using the phyton programming language (Anaconda platform). We aim to find the correlation between user engagement and some of the variables other literature and models (e.g., technology acceptance model) suggested.

The data visualisation elements will be generated using phyton libraries. The data analysis will provide valuable insight into the users' views regarding user engagement and collaboration.

The main risk of this research is the questionnaire response rate. If the response rate is too low, the study can be biased, and the study validity will decrease.

4.2.1. Project Timeline

Here we explain the proposed plan for the research in a broad format. We divided the project into three main stages. During the first stage, the focus is on an intensive literature review and building the questionnaire. The 2nd stage focuse on data collection and analysis; after data collection, two weeks were allocated for the analysis. We aim to collect more than 250 responses, and this may delay this stage. The 3rd stage will focus on the data evaluation and report writing.

We are currently towards the end of the first phase and creating our questionnaire. We hope to publish the complete study of smart libraries at a conference with the smart cities theme in September 2021.

5. CONCLUSION

Smart cities is not a new phrase for developed countries. The digital landscape development and technology revolution has transformed cities into smart cities and all the components. Libraries have an important impact on educating citizens of smart cities and reliantly impact their quality of life. A small proportion of Current libraries are partly automated and fitted with RFID tags and other smart technologies. Most of such libraries are among academic libraries and benefit from being open to the users without being staffed or less staffed than before. Using technologies enables remote control of library buildings, including automatic doors, lighting, self-service kiosks, and public computers. However, it still has a long journey to become autonomous, and many changes are expected.

The influence of internal and external forces is pushing libraries to become more automated and offer alternative services. For example, many new knowledge sources and e-learning platforms have been introduced to the users, replacing the traditional use of libraries. On the other hand, the COVID19 outbreak inforces many sectors and industries for an earlier implementation of automated systems than their predicted time. Libraries were not an exception for this.

These new trends are making the libraries more vulnerable with lower user engagement. This makes the researchers and librarians to re-think of using methods to design the future services that can increase the users' engagement in the next generation of libraries. We aim to find the list of factors that impact the users' engagement in academic libraries. We hope we can use the result to find the correlation between these factors and the users' engagement and present our finding as a framework to help decision-makers design better systems for the next generation of academic libraries.

6. REFERENCES

Butler, K. and Byrd, J., 2016. Research Consultation Assessment: Perceptions of Students and Librarians. *The Journal of Academic Librarianship*, 42(1), pp.83-86.

Chhatwal, A. and Mahajan, P., 2015. CROWDSOURCING IN LIBRARIES: AN OPERATIONAL FRAMEWORK. *International Journal of Digital Library Services*, 5(3), pp.54-58.

Chow, J. (2019). Realising the potential of user surveys for improving academic libraries: The case of St John's College, Cambridge. *Journal of*

Librarianship and Information Science. 51 (3), 689-701.

De, S. and Vijayakumaran, V., 2019. A Brief Study on Enhancing Quality of Enterprise Applications using Design Thinking. *International Journal of Education and Management Engineering*, 9(5), pp.26-38.

Ghosh, P. , Mahesh, T. R. (2015). Smart City: Concept and Challenges. *IJAETS*. 1 (1), 25-27.

Hexa Research. 2017. *Library Management Software Market Size, 2014-2024 | Industry Report*. [online] Available at: https://www.hexaresearch.com/research-report/library-management-software-market [Accessed 23 January 2021].

Howe, J., 2006. The rise of crowdsourcing. Wired Magazine, 14 (6). *URL http://www.wired.com/wired/archive/14.06/crowds.html*. [Accessed 25 January 202]

Islam, A. , Agarwal, N. K., Ikeda, M. (2015). Conceptualizing value co-creation for service innovation in academic libraries. *Business Information Review*. 32 (1), 42-52.

Leorke, D., Wyatt, D. and McQuire, S., 2018. "More than just a library": Public libraries in the 'smart city'. *City, Culture and Society*, 15, pp.37-44.

Neirotti, P., De Marco, A., Cagliano, A., Mangano, G. and Scorrano, F., 2014. Current trends in Smart City initiatives: Some stylised facts. *Cities*, 38, pp.25-36.

Nelson, J., 2016. *Becoming a lean library*. 1st ed.

London: Elsevier Ltd. , 29-37.

Onnasch, L., Wickens, C. D., Li, H., Manzey, D. (2014). Human Performance Consequences of Stages and Levels of Automation: An Integrated Meta-Analysis. *Human Factors*. 56 (3), 476-488.

Ortiz-Fournier, L. V., Marquez, E., Flores, F. R., Rivera-Vazquez, J. C. and Colon, P. A. (2010).

Integrating educational institutions to produce intellectual capital for sustainability in Caguas, Puerto Rico. *Knowledge Management Research & Practice* . 8 (1), 203-215.

Ozeer, A., Sungkur, Y., and Nagowah, S.D. (2019). 'Turning a Traditional Library into a Smart Library', *Amity University Dubai, UAE*, *11-12 December,* 2016, pp. 352-357.

Schöpfel, J. (2018). Smart Libraries. *Infrastructures*. 3 (43), 2-11.

Snehalatha, Veeresh Pujari , Dr. Baswaraj Gadgay. (2018). Smart Library System Using IoT. *International Journal for Research in Applied Science & Engineering Technology*. 6 (7), 471-476.

Tatasciore, M., Bowden, V., Visser, T., Michailovs, S. and Loft, S., 2019. The Benefits and Costs of Low and High Degree of Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 62(6), pp.874-896.

Wojciechowska, M., 2021. Trust as a factor in building cognitive social capital among library workers and users. Implications for library managers. *The Journal of Academic Librarianship*, 47(1), p.102300.