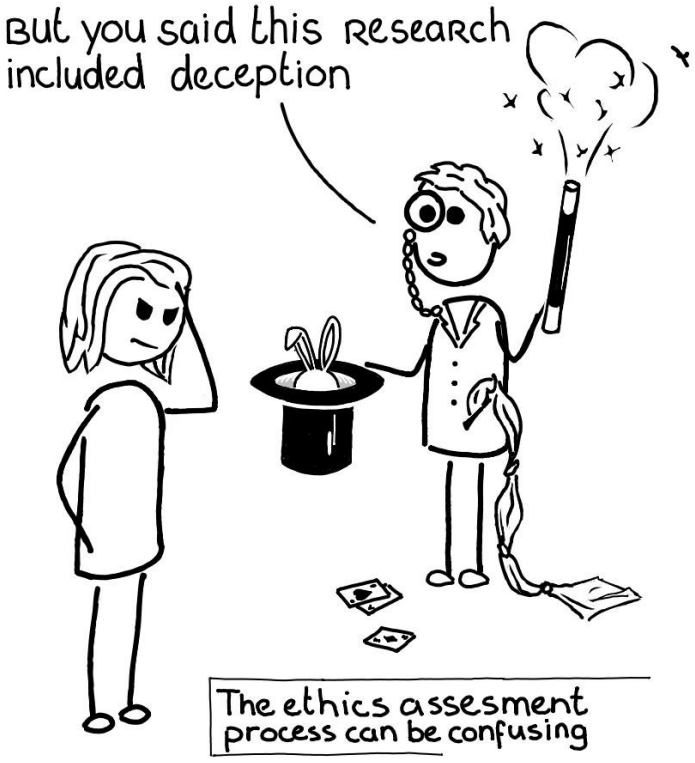
MASTER THESIS

Providing tailored advice for preparing for the ethics assessment procedure

Laura Kester



DEPARTMENT OF HMI

EXAMINATION COMMITTEE
Dr.ir. D. Reidsma
D.P. Davidson PhD.
Dr.ir. R.W. van Delden
Prof.dr. P.A.E. Brey

20-01-2022

UNIVERSITY OF TWENTE.

Acknowledgements

Graduating is never an easy task, even when you really like the subject of your thesis as/which was the case for me. In a long-term individual project, there will always be moments in which it is difficult to keep motivated and keep working towards the final goal. Therefore, I would like to thank everyone around me that I could count on for support, companionship while both quietly working on our own project, motivation to start working in the morning, and distractions from the daily work.

In particular I want to thank my daily supervisor Dennis Reidsma for all the support, fruitful discussions and for being available for feedback or questions. Even when he was extremely busy, I always had the feeling that there was still some time and attention left for me. Next to Dennis, I also want to thank my three other supervisors: Robby van Delden, Daniel Davison and Philip Brey. They were always ready to take the time to provide valuable feedback that lifted my procedures and report to a higher level, even though they heard less often from me. Who also greatly helped me improve my report where the people that reviewed it for me in the final stages. The fresh pairs of eyes I borrowed from Kevin, Jan-Jelle and Thijs have reduced the number of mistakes and have improved the writing and understandability of this report, which I am very thankful for. Lastly, I would like to thank all the people the research done in this thesis could not have existed without, my participants. Everyone who participated in one of the two studies I did during this thesis has provided me with insight in the topic and brought me one step closer to the conclusions I was able to draw.

I am pleased with the result that is lying in front of you, or displayed on your screen. Thank you for all your help.

Abstract

In the Human Media Interaction faculty at the University of Twente there seems to be a lack of engagement in, and quality of, prepared materials for the ethics assessment procedure by both researchers and students. This thesis aims to improve these two aspects during the preparation phase of the ethics assessment process by providing advice tailored to the type of research a researcher or student wants to engage in. To be able to provide fitting advice, the points at which students and beginning researchers struggle most were pinpointed through literature research and research towards practices around ethics assessment in the HMI faculty. These points of struggle were used to provide them with advice that is 1) tailored to the typical patterns of research for their research type and 2) helpful without being too guiding. This led to the development of three versions of advice, the minimum, medium and maximum advice, with different levels of helpfulness and guidance. The minimum advice focuses on short reminders of what not to forget in the ethics assessment preparation, the medium advice also includes examples and explanations about the importance of these points, and additionally, the maximum advice also provides templates and uses a more guiding tone. These three versions of advice were implemented in a newly developed ethics preparation tool in which the information needed to provide tailored advice is gathered through a questionnaire. Evaluation of all three versions with a group of 26 students, inexperienced with the ethics assessment procedure, showed that 1) the medium and maximum advice as implemented in the ethics preparation tool can improve the self-competence of users in going through the ethics assessment procedure and 2) implies that such a tool could improve the quality of the materials prepared for ethics assessment.

Contents

1 Introduction

Ethics and ethics assessment are important parts of the research process in all disciplines. Since the development of the scientific method, researchers have been considering the implications of their experiments in human subject research with varying degrees of concern for possible harm for participants (Ruyter, 2019). However, this process was not formalized until after the second world war (Tribunals, 1949). During the second part of the 20th-century attention for the ethical implications of research only grew, resulting in the adoption of formal ethics procedures in all fields of research (Sims, 2010; Bruckman, 2014). The main goal of these procedures is to protect participants from possible harm caused by research, but they also aim to make the researcher aware of the ethical implications their research might have, and protect research institutions against possible legal actions. Following the ethics procedures can be challenging for researchers for various reasons, causing researchers for example to dread starting the ethics process, have trouble in completing it correctly, or try to avoid (some parts of) it altogether (Nathan, Thieme, Tatar, & Branham, 2016). This can hurt both the speed and quality of research. In the Human Media Interaction faculty (HMI) at the University of Twente (some of) these problems also occur. There is already plenty of research done towards the existence of these phenomena in the field of Human Computer Interaction in general, but practical solutions seem to be lacking.

The topic of this master thesis is to find a way to improve the process of ethics assessment in the HMI faculty by providing adaptive advice to researchers that assists them with preparing for the ethics process, while ensuring they retain a sense of ownership of the materials prepared.

To achieve this, a clear understanding of the situation around ethics assessment and challenges that occur in the field of HCI, to which the HMI faculty belongs, is gained as a theoretical basis through a literature review. In this review, the history of ethics assessment and the current situation in HCI are explored. The insights gained are used to formulate the problem statement and research question for this thesis. Then, the insights from the literature are compared to the situation at the HMI faculty at the University of Twente through research in and around the HMI faculty towards the practices around ethics assessment, during which researchers, students, and ethics committee members share their views about the current state of the ethics assessment process. Based on the information gained from literature and research towards practices around ethics assessment in HMI, three different types of advice questionnaires are developed inside a pre-existing adaptable advice tool. In the last part of this thesis, these versions of advice are compared based on five different measures related to helpfulness, effectiveness, and ownership to evaluate to what extent they are able to improve the process of preparing for the ethics assessment in HMI.

2 Background and literature review

2.1 History of research ethics

The word ethics can have slightly different meanings in different contexts. In this review, we will go with a definition that seems to be close to the definitions used in the literature from the field of HCI. Ethics in this context is the discipline of philosophical and systematic study of morality where moral concepts such as good, bad, responsibility, etc. are analyzed (Frauenberger, Rauhala, & Fitzpatrick, 2016; van der Burg & van de Poel, 2005). In a research context, (fundamental) concepts in ethics are applied to research methods, which results in good practices about how to conduct research ethically (Frauenberger et al., 2016; Luger & Rodden, 2013). Research ethics does not concern the professional ethics, or professional behavior, of the researcher, but only the ethics of the way research is done. So for example, topics within research ethics are respect towards society and others, the use of resources and research outputs, and ideas about good scientific conduct. However, before looking into research ethics in the field of HCI, the history of research ethics is studied to get an idea about its origins.

Research ethics originate from the practice of medicine and some form of ethics codes already existed in the time of the ancient Greeks (Ruyter, 2019). The most famous example of this is the Hippocratic oath. This oath was used as a guideline on how to be a "good" physician. Even nowadays an oath loosely based on the Hippocratic oath is sworn by medical students before they receive their medical license (KPMG, 2017). Furthermore, many philosophers have published about ethical behavior in scientific research, like Francis Bacon who argued that scientific research should always benefit humanity in his work on the scientific method (Scharff & Dusek, 2013). However, these early codes of ethics were self-regulated by the profession and often implicit (Ruyter, 2019).

2.1.1 The Nuremberg code

It was not until the end of the Second World War that an ethics code for (medical) research called: "The Nuremberg Code" was created as part of the conviction of 23 Nazi doctors(Tribunals, 1949). This code was created to make sure that atrocities such as they had committed would be prevented. The code consisted of 10 rules which define the “basic principles that must be observed to satisfy moral, ethical and legal concepts”. It was later endorsed by the World Medical Organisation with the Declaration of Helsinki. Nevertheless, governments and other organizations still conducted some questionable experiments after 1949, even though there was a basis on which to question and condemn such experiments (Ruyter, 2019).

2.1.2 In the U.S.: the Belmont Report

Because a significant part of the literature about ethics in HCI was written by authors with a US perspective, the history of research ethics in the US also has some importance. In the United States, one especially cruel case of unethical experimentation on human subjects, the Tuskegee syphilis study, has had a great impact on research ethics. In this study participants with syphilis were unaware of the study and intentionally left untreated, even though an affordable and effective cure was developed while the study was still taking place (Sims, 2010; Bruckman, 2014). The U.S. National Commission for the Protection of Human Services of Biomedical and Behavioural Research reacted to this with the Belmont Report. This report, which identifies basic ethical principles regarding research involving human subjects, has led to improved regulations regarding studies with human participants. It outlines three core principles that need to be respected when conducting research with human participants. These are (1) beneficence, (2) justice, and (3) respect for persons (Sims, 2010; Bruckman, 2014). The first principle of beneficence is often interpreted as "do no harm", but this is too straightforward. Rather it means that the possible benefits must outweigh the risks, which is sometimes difficult to

determine since benefits and risks often fall on different individuals. Here the principle of justice comes into play. Justice in this context means that risks and benefits should be equally distributed over society. Equal distribution of risks is quite difficult to realize if you think about it since only a limited number of participants are taking part in studies. Next to this, including vulnerable groups in society who may also benefit can lead to extra work and precautions and is therefore often neglected. The third principle of respect for persons tells us to treat people as ends themselves, and not as means for an end. This principle is somewhat similar to the principle of justice but on a more personal level. Every individual should be treated in a way that their autonomy and needs are respected. As a result of the Belmont report, institutional review boards (IRB) were developed which even have legislative power (Bruckman, 2014). These review boards are tasked with overseeing that all research is done within the bounds of the three core principles. However, the three Belmont principles are quite broad, so many institutions have clarified or expanded on these points in their own ethics codes.

2.1.3 In the Netherlands: the declaration of Helsinki

In the Netherlands, just as everywhere else, there have been scandals regarding questionable experiments where participants were not aware of the risks, but not of such a scale as the Tuskegee study. Therefore, there was no large public push for the definition of ethical principles beyond the declaration of Helsinki made by the World Medical Association (WMA) (Goodyear, Krleza-Jeric, & Lemmens, 2007). Nevertheless, ethics committees such as the American IRBs did develop. Not as a reaction to past abuses, but as a precaution against possible abuses. However, during the second part of the 20th century, the problem of possible abuses was still only regarded as a professional medical problem. The council for public health (Gezondheidsraad) did demand action on this topic from the legislature (Tweede Kamer) multiple times, but it took until 1998 for any law regarding this subject to be put in place (*Wet op het bevolkingsonderzoek - BWBR0005699*, 1998).

2.1.4 New field of study: Human Computer Interaction

Originally, committees reviewing research ethics were set up for medical and psychological research, but as other research domains developed and started researching with human participants, the existing policies and procedures were translated to these new domains. However, not all of them translated easily into these new domains (Bruckman, 2014). This was also the case for HCI. In the beginning translation of procedures to HCI was challenging because of a lack of knowledge about computer systems etc. on the side of the IRB members, who originated mostly from the medical and social sciences. However, this has improved significantly over time due to the diversification of IRBs with members from different disciplines, and because IRB members themselves got more familiar with the new subjects of study. However, the field of HCI has continued to evolve. Where the optimization of usability might be a good enough principle in the human factors paradigm, which was very prominent in the early days of HCI, this is not sufficient anymore (Frauenberger et al., 2016; Nathan et al., 2016; Munteanu, Waycott, & McNaney, 2020). Because technology becomes more and more embedded in our daily lives the values embedded in technologies and how society interacts with them become more and more important. So, the situated practices paradigm in HCI has become more prominent in recent years. This paradigm has a focus beyond mere function and performance regarding design, giving room to the participation of stakeholders in the design process. Furthermore, the introduction and rapid growth of the internet has brought about new challenges regarding consent, researching outside of a lab setting, and other features of the ethics process. All these developments gave rise to new challenges regarding the formal ethics process in HCI, which will be explored in the following chapters.

2.2 Ethical principles in HCI

To be able to understand the situation around the formal ethics process in HCI, it is crucial to understand what ethical principles this process is based on. The ethical principles in HCI are partly general principles that apply to all fields of research, and partly principles that are specific to the broader field of computer information sciences. According to an extensive review of different formal and informal sources done during the SATORI project, 8 Ethical principles apply to all fields of research and innovation (Jansen et al., 2016).

Research integrity. When conducting research, researchers are expected to behave appropriately, behave respectfully towards participants and give proper acknowledgment to sources and others who contributed to research.
Social responsibility. Researchers should always keep the benefit of society in mind. Especially if their work is funded with public money.
Protection of and respect for human research participants. If participants are involved, they should always be treated with respect and human decency. Participation should always be voluntary and participants have to be protected from harm as much as possible.
Protection of and respect for animals used in research. Animals should always be treated humanely.
Protection and management of data. Researchers are expected to protect participant data and manage it in such a way that the privacy of participants is not harmed in any way.
Dissemination of research results. Results have to be reported. This has to be done in such a way that they are unambiguous and replicable. If done well, another scientist should be able to repeat the experiment and get similar results.
Protection of researchers and research environment. Researchers and research environments should not be endangered by conducting experiments.
Avoidance of and openness about potential conflicts of interest. (The appearance of) a conflict of interest should be avoided to make sure results are unbiased.

According to this same review, there are also at least four ethical principles that apply specifically to the field of computer information sciences, and thus HCI. These field-specific principles depend on the characteristics of the field and the types of research generally conducted.
Avoidance of information and security risks. Fields like HCI often make use of large amounts of data and/or conduct internet research, so data security becomes an important ethical consideration in both research and development.
Respect for privacy and personal data. Not only is it important that information is secure, but also that when results are published or a data set is shared, the privacy of participants or users is protected.
Social responsibility in the sense of biases in new technology. New developments in technology can have implications for social relations and can harbor unjust biases. It is therefore the researcher's social responsibility to ensure that possible new technologies include respect for freedom of expression and intellectual property rights, and do not harbor or counter unjust bias in terms of age, gender, sexual orientation, social class, race, ethnicity, religion or disability.
Dual use. In the computer information technology field there is a reasonable chance that new technologies in computer information sciences have a dual-use, for example, use in a military application. The implications of this on ethical research towards this new technology have to be considered carefully, to be aware of and possibly adjust the development of technology with very questionable dual-use applications.

These field-specific principles, combined with the more general principles form the basis of the ethical guidelines in the field of HCI.

2.2.1 Focus on principles of informed consent and privacy

Two of the most influential principles for dilemmas in the research ethics in HCI seem to be the principles of privacy and informed consent, due to their explicit presence in ethical research practices. So, they will be discussed in more detail in this section. Their origins, however, lie in quite different directions.

Principle of informed consent

Because concepts of research ethics in HCI have their origin in medical and social sciences, in which informed consent plays a very important role, informed consent got an important place in research ethics in all fields of study, and thus also in HCI. Informed consent means that participants are participating freely in the study, that they are aware of the nature of the study and possible risks associated with it, and that it is offered in a fully understandable form. Given some of the examples, like the syphilis study in the previous chapter, it is clear why informed consent is so important from a medical sciences perspective. If the participant is not aware of the risks and goal of a study, they will not be able to protect themselves against unreasonable harm.

Since ethics in medical and social sciences are greatly influenced by the Nuremberg code and declaration of Helsinki, contemporary concepts of consent in many different fields of research are too. In these two codes, informed consent is based on 4 principles. Consent has to be voluntary, competent, informed, and comprehending (Luger & Rodden, 2013). The goal of informed consent is to make the participant in a study aware of how a system functions and support them in understanding what possible risks of participating are, so they can make an informed choice about participation (Luger & Rodden, 2013; Brown, Weilenmann, McMillan, & Lampinen, 2016).

Next to the research ethics perspective, consent is also important nowadays because of the rules and regulations around privacy. These rules and regulations focus on the data protection of individuals. According to the European GDPR, a user or subject needs to give informed consent for the processing of any personal information and has the right to access and deletion of this data at any time. Information about the processing of data and the user's rights have to be presented "in a concise, transparent, intelligible and easily accessible form, using clear and plain language" (Regulation, 2016). Both perspectives of consent must be taken into account when asking a participant for informed consent.

Principle of privacy

The principle of privacy does not originate from the medical and social sciences and its importance is a relatively recent development. Up to a few decades ago privacy was not something researchers and participants had much concern about. However, changes in the possibilities of data collection during research and the ease with which data can be gathered, processed, and shared through modern technology, have changed research practices and increased the perceived importance of privacy. This growing importance of privacy in recent years is also one of the reasons the GDPR, mentioned in the previous section, was put into place.

In a research domain such as HCI where modern technology is both a research topic and way of data collection, the principle of privacy plays a large role in research ethics nowadays. Almost all research in HCI includes the recording, processing, saving, and/or publishing of data. Data can consist of answers to a questionnaire but also things like recordings, phone usage, sensor data, etc. Some data can be used to identify a person, even if this data is not linked to a name or address. Exposing this data can harm participants' (sense of) privacy and security.

Even though privacy and protection of personal data is an important consideration for researchers and participants, there is no clear consensus about what that privacy exactly entails. According to Kevin Macnish the accounts of *privacy as control* and as *privacy as access* of information are most widely used definitions of privacy (Macnish, 2018). The *privacy as control* account states that privacy is lost

if control of personal information is lost. The *privacy as access* view states that there is only privacy lost if the information is accessed by someone. There is no clear consensus within the scientific community on which definition is right. Both are therefore often used in discussions about privacy and privacy loss, sometimes interchangeably. However, not only the actual loss of privacy is important but also the participant's perception about their privacy. Even if no loss of privacy occurs the participant can still experience harm because of a perceived loss of privacy (Macnish, 2018).

This distinction between a perceived loss of privacy and a factual loss of privacy is clarified by Moore (A. D. Moore, 2008). Moore makes the distinction between privacy as a *right* and privacy as a *condition*. A perceived loss of privacy is a loss of the *condition* of privacy, while an actual loss of privacy is a violation of the *right* to privacy. If one is in a condition of privacy, they will be assured that no one can invade their privacy unnoticed. According to Moore, this can be compared to being naked in a room with no windows, so you are certain no one is invading your privacy. If one is not in a condition of privacy, it does not mean that one's right to privacy is also violated. For example, if someone placed a camera in the room, your right to privacy is not violated if no one ever sees the video footage of you naked. However, you will probably not experience a feeling of privacy because of the camera being there. Therefore, you are not in a condition of privacy anymore even though your right to privacy might not be violated. This distinction is easily overlooked, even though it can make quite a difference for the harm done to the person whose privacy might be invaded. Therefore, it will be addressed explicitly when privacy is discussed.

2.3 Dilemmas in HCI regarding principles in research ethics

It is not possible to determine a comprehensive set of research practices that fit all the principles defined in the previous section, because some principles lead to contradicting outcomes when put in practice, or influence the validity of results. When the focus is put on one principle, this could create tension with another principle. How ethical a certain research practice is, is greatly influenced by which ethical principles are deemed most important. In this section, existing tensions between ethical principles in HCI and the tensions between ethical principles and research outcomes are discussed.

2.3.1 Dilemmas regarding inclusion and diversity

In all fields of research, it is of great importance for inclusivity and equality that participants from different populations are included as part of general research since research should benefit the whole of society. However, there exists a tension between the inclusion of diverse participants, and the principle of informed consent when doing research with vulnerable groups. Sometimes, putting emphasis on one principle can inherently exclude the other (Sanderson, 2010; Sudore et al., 2006). For example, when medical research is conducted, kids are often excluded. After all, they cannot give informed consent themselves, because they are minors. This results in medication not being approved for kids, even though some of them might benefit greatly from it. In HCI this dilemma manifests itself in the consideration between viewing a certain group as an anomaly and excluding them because it would be difficult to get informed consent, and trying to include them to make sure a technology benefits all even though the validity of informed consent might be questionable.

Furthermore, there exists a tension between the diversity (benefit for all) and the quality of research outcomes. This tension is caused by the benefit of uniform sampling to the quality of research (Abascal & Azevedo, 2007). Doing research with a uniform sample of participants makes it easier to get unambiguous results, but is at the expense of participant inclusivity, resulting in reduced equality in scientific research. Even though this tension exists in all fields, it is especially visible in HCI, because a lot of research in this field aims to change the world for the better, often doing research or designing for situations with great societal tension (design 4 good). Furthermore, technol-

ogy developed in HCI is often aimed to be widely used which makes errors regarding inclusivity more visible, which is less the case in other fields of study. For example, compare testing the effectiveness of medication with testing the effectiveness of a facial recognition program. If both technologies are tested with a participant group existing of only Caucasian men, both technologies are possibly racially and/or gender-biased. However, biases will probably be more obvious for technologies used in HCI, such as facial recognition, because of a confrontation with these biases due to their intended use. Users from different racial and/or gender groups are confronted more directly if they are identified wrongly based on their facial features, compared to when they use medication that might not be as effective as it could be. Therefore, inclusivity in the sense of not excluding groups from general research gets special attention in HCI, even if it is a problem experienced in all fields of research.

2.3.2 Dilemmas regarding the principle of informed consent

Some research practices used in HCI have a higher level of unpredictability and a shift of knowledge about risks from researcher to participants. This is of course not always the case in HCI research. It is still common for participants to not have substantial knowledge to assess risks, and that the contents and course of the experiments are known. But if these other practices are used, it is difficult for the researcher to get fully informed consent from the participant beforehand. An example of this is the participatory design method in which the participant also engages in the design of a tool or application. During this process, the roles of the researcher as expert and participant as test subject become blurred, and in some cases, participants might know more about some parts of the topic than the researcher. Also, it can become clear during the experiment that the needs of the participant differ greatly from what the researcher predicted, making it difficult to determine the end goal of the experiment and get informed consent for participation. These kinds of complications during research are illustrated well in Branco's paper about co-design with dementia patients and their families (Branco, Quental, & Ribeiro, 2017). The persons close to the patient are more capable to assess what is potentially harmful to the patient than a researcher is, making them more of an expert in certain parts of the study than the researchers themselves, which complicates an otherwise relatively straightforward process of anticipatory consent.

A dilemma with the principle of informed consent can also arise when research does not carry (serious) risks for a participant's mental and physical health. In those cases, getting informed consent according to the rules and regulations in place can do more harm than good. Of course, if there are substantial risks in participating, elaborate and strict consent practices are a valuable tool to make sure participants are informed and participating voluntarily (Wendler, 1996). There is quite a lot of HCI research that carries significant risks such as the deception of participants. However, some typical research topics, such as user experience design, do not carry large risks (Brown et al., 2016). The regulations for experiments with minimal risks still require specific actions such as ticking a box or verbally assenting to certain statements (Luger & Rodden, 2013; Brown et al., 2016). This can take up significant time, can be perceived as boring and unnecessary (Brown et al., 2016). Also, asking for consent can devalue results when part of the experiment is about testing something in a natural environment. In cases like this, the participant being aware of the experiment before or during the experiment ruins the experiment itself (Waern, 2016). Furthermore, participants do not seem to care that much about consent practices in these types of situations. Often, participants have no recollection of what they agreed on or what was written in a consent form at all after the study and do not even seem to want to read the form (Brown et al., 2016; Luger & Rodden, 2013). So, they just agree to everything without reading. In these cases, the goal of informing the participant was not achieved by the consent practices in place. The only goal that might be achieved is compliance with the GDPR by the researcher. Solutions to accommodate for experiments with minimal risk are already proposed by the community, although they are not applicable in the short term. For example, there is a call for more discussion on ethics and consent around low-risk research within the community, and principles

to re-frame consent practices have been formulated (Brown et al., 2016; Luger & Rodden, 2013). These developments could in the long term improve consent practices in HCI.

2.3.3 Dilemmas regarding the principle of privacy

The principle of privacy states that the privacy of the participant must be protected to minimize possible harm. However, the resulting practice of fully anonymizing data is in tension with the quality of research outcomes, because in the process of anonymization valuable information could be lost. To get a clear idea of why this is the case, we first have to understand why and how data is anonymized. In general, personal data collected during research is anonymized to protect participants. Data is seen as personal if the data allows for the person the data originates from to be identified (European Commision, 2018). The use of personal data is also protected by law by the European Commission so that it can only be processed if consent is given by the person the data originates from. To make data anonymous, it is not enough to just remove a person's name and address. Depending on the type of information gathered participants can still be identified based on a combination of different information that seems unimportant by itself (European Commision, 2018). For example, someone's study program, amount of social media use, and favorite music might be enough to determine someone's identity via social media. This causes researchers to remove or generalize data points to such an extent that the value of the data is limited, making them choose between the anonymity of participants and the quality of research. Therefore, anonymization limits the researcher's ability to insight from the data, but it is still done in the interest of the participant.

However, in some cases, a participant may be able to benefit more from non-anonymous participation in research. Since this goes against the principle of privacy, it is generally not possible, even though participants might benefit from being non-anonymous for multiple reasons (Brown et al., 2016). First of all, participating in research that can have great benefits to society can also be seen as something worth sharing. Especially when data is not sensitive and risks of research are low, which is common in HCI (Bruckman, 2014). Next to that, when a participant was part of the co-design of a new system or technology, they might be proud of their achievements and would like an acknowledgment of their efforts. It could be interesting to approach privacy in HCI not as something to be protected at all costs, but as something that can be discussed freely with a participant in these kinds of cases (Brown et al., 2016). However, this may not be possible while protecting a participant's privacy.

2.3.4 Tension between protection of human participants and use of public data

In all fields of research, researchers are expected to behave respectfully towards participants, both during experiments and in their research design. In HCI, there is an ongoing discussion about if the method of using public personal data for research is a research design that is respectful towards persons involved. Although this discussion is relevant for all fields in which public data is collected, it is especially relevant for fields in which internet research is a widely used research methodology, like HCI. An example of such a questionable method is data from social media sites collected without the knowledge of the user. Technically, this type of research does not have participants. Users give away rights to the technology in exchange for other services, such as being able to connect with others free of charge (Calvo, Peters, Huppert, & Goggin, 2018). Data is collected by researchers independently. However, they do consent to the possibility of this happening when they start using this service. What they may not be aware of is that these kinds of services are notorious for changing their privacy setting and making them difficult to comprehend (Cooper & Coetzee, 2020). They will most likely also not notice if the same data is collected by a web crawler and used for research. Even though policies of social media sites clearly state to users that this is a possibility if they make their information public, it is questionable if users have read these policies and are aware of this possibility (Moreno, Goniu, Moreno, & Diekema, 2013). It can therefore be argued that they were not able to give consent. On

the other hand, it still is the responsibility of the user to be aware of what they agree to. This creates a moral dilemma for researchers who want to use this kind of online data. They are not violating any law in collecting this kind of data; however, it may still feel instinctively wrong to do so depending on how the data is used.

If one would follow the *privacy as access* account, using public data is perfectly legitimate (Macnish, 2018). In the access account of privacy, the condition of privacy is neglected. Privacy is only violated if data is accessed which the user did not consent to be accessed. The user has consented to the data being public and thus, accessible to everyone. Therefore, using this data for research will not be seen as a privacy breach, even if the data is processed and published in a paper. However, if one looks at *privacy as control*, using this kind of public data for research becomes slightly more problematic. (A. D. Moore, 2008) On a social media platform, the user can control which data is public and not, and change their preferences in this regard. When data is collected from such a platform the user loses this control and has no influence on how the data is processed and published. They can decide to make the data on the social media platform private, but they cannot do the same for the data collected earlier by a researcher. Thus, the user experiences a loss of control and privacy. The loss of control results in the person in question losing their condition of privacy, even if the right to privacy is not actually violated during the study. Of course, it also depends on the type of research and type of data collected if people would experience this as problematic. For example, analyzing people's tweets to correlate their amount of grammatical errors to the political opinion they share on social media, will be perceived as more problematic than using the same data set to correlate the difference in vocabulary with age. Depending on which view on privacy one finds most convincing, the source of the collected data, and the type of research; a researcher may feel morally obliged to ask consent in some form for using public data during internet research.

2.3.5 Tension between avoidance of security risks and generation of results

In the field of information sciences generally, third-party technologies and tools are used to collect data for research. These may have their own data collection and privacy policies that are important to consider when selecting tools for data collection because some of them might not prioritize the interests of the user. Examples of this are lifestyle apps, screen recorders, and activity trackers. These tools can record a wide variety of data such as phone usage, steps, sleep patterns, etc, and have the benefit of using the users' smartphone as a sensor. Since almost everyone has a smartphone, no new sensors must be bought by the researcher, or worn by the participant; making data collection cheaper, easier, non-invasive, and durable. When a participant is asked to use an app for the duration of research activities, they also have to agree to the terms of service of this tool. Some tools might store information on a server outside Europe where the laws around privacy are different or might compile the data themselves to gain insight into the user population. Ideally, these are also considered by the researcher in selecting the app. However, reviewing the policies of a third-party technology is not part of the formal ethics assessment procedure and can easily be forgotten. Even if a researcher has reviewed the third parties' policies and makes participants aware of these, participants can forget to discontinue tools or permissions and have their data collected over a long period of time without their knowledge. Because of the clear benefits to data collection, the use of third-party tools for data collection should not be prohibited. However, the ethical implications of using third-party tools can be an important consideration for the researcher and, if necessary, should be brought to the attention of the participant.

2.4 The ethics assessment process in HCI

Based on the principles from the previous section, a process of ethics assessment of research is adopted by the scientific community. In HCI, as in many other new fields of study, this process is

based on existing processes in medical or social sciences. Typically, the process of ethics assessment in HCI consists of some form of self-assessment, the preparation of materials such as the consent form, and the submission of a proposal to an ethics committee. Based on the self-assessment and materials prepared, the committee will approve or decline your proposal based on the institute's guidelines. Even though these guidelines are all based on the same set of principles, the focus they choose can differ per institute; and thus dilemmas that occur can differ as well. This is not only influenced by the field of study or type of research, but also on contemporary themes committees might deem more important. When a proposal is declined, one has to improve or reformulate their proposal so it adheres to said guidelines. A researcher needs to acquire approval from the ethics committee before they start their research.

2.4.1 Factors that hinder HCI researchers engagement with the process

HCI researchers often seem to have some resistance regarding going through the formal ethics process. Many of them quickly learn to say the right things to an ethics committee to get approval with minimal effort (Guillemin & Gillam, 2004). Sometimes, they even deceive review boards intentionally to avoid them (Davis, Whitfield, & Anwar, 2018; Vitak, Proferes, Shilton, & Ashktorab, 2017). Even if researchers do not have this resistance they experience difficulty in getting ethics approval. There are multiple possible causes for this already documented in the literature.

Guidelines are hard to find and/or apply

The guidelines can be difficult to understand and follow properly. This is not a problem unique to HCI since researchers in many disciplines struggle with these in some way. However, the specific challenges do differ. In literature the following challenges are mentioned regarding HCI :

- The guidelines are written in a language that is hard to understand (Molich, Laurel, Snyder, Quesenbery, & Wilson, 2001)
- The guidelines are difficult to locate (Molich et al., 2001)
- Standards and regulations are sometimes inconsistent and too general (Davis et al., 2018; Vitak et al., 2017)
- There are few (if any) examples of how these principles are applied in practice (Molich et al., 2001)
- Some of the guidelines mostly deal with usability testing and handling of videotapes (Molich et al., 2001)

All of this can make it challenging for researchers to find out which guidelines apply to their research, how to properly prepare for an ethics review by the review board; and discourages some researchers to start the process.

Guidelines do not promote working with vulnerable target groups

Guidelines in HCI sometimes unintentionally disadvantage vulnerable groups. The main reason for this is that guidelines about these groups are sometimes overly complicated. Special regulations are in place to protect vulnerable populations from harm when research is conducted with them since it can be problematic to get informed consent from people belonging to a vulnerable group.[1] This is in principle a very good thing because it is undesirable if people who already have a disadvantage in society are exposed to even more possible harm without giving proper informed consent for this. Especially given the history of medical experimentation on children, the poor and disadvantaged in mind (Resnik, 2021). But, these guidelines can get overly complicated and discourage researchers

[1] see 2.3.1 for more explanation on this topic

from both including vulnerable groups in general research, and focusing on research for them specifically (Brown et al., 2016; Nathan et al., 2016).

Insufficient skills of researchers in the preparation of materials

Researchers in HCI often lack the expertise to adhere to formal ethics protocols, which is at least partly due to their complicated nature and lack of examples of how to apply them (Molich et al., 2001). So, even if researchers find the right guidelines to follow for their type of research, they are often not able to follow them correctly. Many bachelor's and master's programs at universities already try to improve these skills by teaching students how to properly engage in deliberation about ethical decision-making. However, this does not seem to always come easily to researchers in technical fields such as HCI (van der Burg & van de Poel, 2005). Ethical decision-making does not seem to be part of the culture in these study programs, which implies that, when these students become young researchers there is a significant chance that they lack the tools to engage in ethics properly. This makes it more difficult for them to go through the formal ethics process.

Since many researchers already struggle to go through the regular ethics protocol, they are even less likely to be able to deal with the more complicated guidelines for research on vulnerable groups (Munteanu et al., 2015). For researchers, it can be such a burden to deal with all regulations around including vulnerable individuals that it is easier not to include them at all. This can in turn lead to skewed results and fewer opportunities for vulnerable groups to benefit from modern technology.

Review boards are sometimes contra-helpful

Researchers and guidelines are not the only problems. The people who have to enforce guidelines, the review boards, are often understaffed and overworked (Vitak et al., 2017). The fact that many researchers have to go through the approval process multiple times only adds to this. Also, the members of a review board do not always have a consensus about what should be reviewed, which is probably influenced by the confusing guidelines (Davis et al., 2018; Vitak et al., 2017). The combination of review boards being short on time with the fact that members are not in agreement about the review process makes that boards are probably not able to look closely at the details of an application and might disapprove more risky research unnecessarily.

Furthermore, this extra careful behavior can also add to the exclusion of vulnerable groups because including vulnerable populations is perceived as an extra risk (Brown et al., 2016; Luger & Rodden, 2013; Munteanu et al., 2020). Even if a researcher decides to explicitly research with vulnerable groups, which is already difficult due to the tension between inclusion and consent discussed in 2.3.1, review boards are extra careful and may hesitate to grant approval for the proposed research.

Lastly, ethics review processes suffer from ethics creep leading to review boards applying guidelines increasingly strict over time. The ethics creep is characterized by a dual process in which the regulatory system is expanding new activities and institutions, while at the same time intensifying the regulation of existing activities (Haggerty, 2004). This aggravates existing problems causing a vicious cycle of more and more complicated processes and restrictive regulations, which in turn can lead to researchers feeling censored by review boards and slow down research. Ethics creep can cause the system of ethics review to do more harm than good in some cases, since the harm done by slowing down research can become greater than the harm to participants that is prevented by review boards (Brown et al., 2016).

HCI is exceedingly situational while anticipatory processes are not adjusted to this

The fact that the ethics assessment process in HCI is mainly anticipatory leaves large parts of the process unaddressed, because formal ethics procedures often fail to give researchers appropriate guidance in dealing with ethical implications of research methods or unpredictable environments (Frauenberger

et al., 2016; Nathan et al., 2016). This is partly because the type of research practices used in HCI do not fit the anticipatory nature of the current ethics process and the different roles researchers have, as already discussed in 2.3.2. Because of the lack of guidance, researchers are left to deal with these issues themselves.

Because HCI research often takes place outside of a standard lab setting, it is impossible to predict all possible risks, burdens, and harms. This makes it difficult to address ethical challenges beforehand, which is required in the current ethics assessment process. As already addressed in 2.3.2, participants are more involved in the process instead of being experimented on. And, since the group of participants is often quite diverse, reactions in certain situations are diverse as well. Not every event that can take place in a more or less real-life setting can be known. Therefore, anticipatory ethics procedures are not able to address ethical dilemmas arising in design-based explorations or other unpredictable situations (Frauenberger et al., 2016; Brown et al., 2016; Nathan et al., 2016).

Existing procedures do not provide assistance in dealing with situational ethics after approval

These unexpected ethical challenges require the researcher to make ethical decisions on the fly (Munteanu et al., 2015; Frauenberger et al., 2016; Nathan et al., 2016). However, researchers in HCI are often not prepared for this. The anticipatory ethics assessment procedure they go through does not provide support for dealing with risks and harms in an unpredictable situation, because they cannot all be anticipated. This makes that all responsibility for ensuring ethical research falls back on the researcher (Guillemin & Gillam, 2004). What can complicate this, even more, is that in research procedures used in HCI like co-design or long-term exploratory studies, researchers sometimes get quite close to participants' lives which makes some ethical challenges more personal and difficult to resolve (Munteanu et al., 2015; Waycott et al., 2017). If there was a culture in place within HCI that allows for arguing about ethical considerations back and forth before and during a study, some of the issues regarding the unpredictability of research may be avoided. Such a culture could be the glue between formal processes and the unpredictability of different environments (Munteanu et al., 2015; Brown et al., 2016). However, as implied in 2.4.1, this is not the case and ethical issues remain unaddressed or are dealt with in the wrong way. This can not only unintentionally harm participants, but also researchers.

A good example of this is given by Waycott. When a researcher is observing the use of some software application in an adult reading class over a long period of time they will get to know their participants quite well and may unintentionally develop a bond. Participants might start to view the researcher as a friend or caregiver. If something troubling happens in the participant's life the researcher is also confronted with that and the participant may look towards the researcher for help or advice. In social and medical sciences researchers are taught how they could handle these kinds of situations, but that is not the case for HCI. Since they do not know how to deal with these situations they may not know what is the right choice to make to resolve the situation and when to keep their distance. This can put emotional strain on both researchers and participants.

2.4.2 The need for Improving ethics assessment in HCI

Even though some projects have already been set up to improve ethics assessment, it is not enough. Findings from the SATORI project imply that tools to improve ethics assessment in the information technologies domain are lacking (Danish Board of Technology, 2015). SATORI's detailed analysis on what efforts are already being done to improve ethics assessment shows that training activities on the subject of ethics and other resources originated from the medical and social sciences and support for developing countries. Also, there are few resources available that assist or train ethics assessors in assessing new technologies.

However, the few projects and conferences that do talk about improving ethics assessment are quite

insightful even if they do not provide concrete solutions. In the field of HCI, there have been several conferences and workshops about the topic of ethics assessment, which make valuable contributions towards identifying the problems in the ethics assessment. Only some recommendations were provided towards a more concrete solution in the SATORI project and the Agora project (van der Burg & van de Poel, 2005; Danish Board of Technology, 2015). They provide suggestions on what young researchers would need in a possible training program. These include opportunities for multidisciplinary discussions and participatory processes on the role of science in society, gaining knowledge of the basis of ethics assessment, getting tools to identify ethical issues, and evaluating their research design. Even though this information is limited, it provides some direction towards improving ethics assessment practices in HCI. In this thesis project, we will focus on one area of such improvements.

3 Problem statement

Both training/assistance in ethics assessment and more comprehensive rules and regulations for the formal ethics assessment process could help to improve the situation around the formal ethics procedure in HCI. Stakeholders, researchers, and review boards seem very willing to work on improving this situation. Sadly, the codes of ethics currently available are not likely to change anytime soon. Even if it would be decided that some changes should be made this is most likely a long-term plan, due to the complicated nature of guidelines and the number of stakeholders involved. However, improvements of the ethics process can also be achieved within the current codes of ethics available by providing more clarity, less ambiguity, improving understanding regarding the ethics process, and providing helpful and tailored advice to researchers as they are engaging with the EC process to deliver higher quality materials. This master thesis will therefore focus on training and assistance in ethics assessment, to improve the researcher's skills in ethics assessment.

However, even though it is clear from the literature that many researchers struggle with the process and guidelines, it is still unclear at what specific point in the process researchers struggle most and what kind of help they would need to make these parts easier to get through. Ideally, any form of guidance through the formal ethics process should also strengthen their ethics deliberation skills, to make them more capable of making ethical decisions themselves. Especially because in HCI not all ethical challenges can be addressed in the formal ethics process. Furthermore, it is unclear what would be the best format in which to provide such assistance. Therefore, this master thesis will explore the problematic points of the ethics process in detail to be able to determine how to provide the right information at the right time through an already existing adaptive advice tool (described in future section 5). Of course, the effectiveness of the adapted tool will be determined through evaluation.

3.1 Research questions

From the research objective, three main research questions can be distilled.

- **RQ1 –** At which specific points in the process of gaining ethics approval do researchers struggle and why?
- **RQ2 –** How can support on these specific points be implemented in the available tool while balancing helpfulness to the researcher, and the amount of learning and ownership of the researcher?
- **RQ3 –** What is the effectiveness and impact of the resulting tool?

To answer the first question research has to be conducted at the HMI faculty to get a deeper understanding of the problem by identifying the context around the formal ethics process at HMI specifically, and identifying points at which researchers and students struggle and what support they would like.

If points of struggle points are clear, the second question can be answered by implementing the insights gained from the first question as a tool that might provide the necessary support. At the moment a very basic tool is already made which allows researchers to prepare their proposal for the ethics committee by answering a set of questions. Depending on some of the answers given, the set of questions is changed so that researchers are only answering the questions that apply to their proposal. If they finish these questions they are presented with complete documentation they need to properly complete the request to the ethics board. This tool can be used as a starting point for answering the second question, which involves fine-tuning of questions and the format of the advice that is delivered.

Lastly, the third question can be answered by evaluating the effectiveness of the resulting tool through user testing it with participants who are asked to prepare for a fictional ethics assessment using the advice from the tool. Ideally, the resulting tool would both provide support for researchers and improve their skills of ethics deliberation.

This master thesis, given all questions are answered satisfactorily, will not only result in new insights about the ethics assessment process in HCI but also on a new tool that provides support for researchers going through the ethics process while improving their skills.

4 Research towards practices around ethics assessment in HMI

In this section, the method, execution, and results of the research towards the practices around ethics in and around the HMI faculty are discussed and compared to the literature research. The literature research in the previous section only provides a general image of the problems that occur during the ethics assessment procedure. Also, the papers available are based on the ethics procedures in different institutions operating in different countries, which makes it unclear if they accurately describe the situation at the University of Twente. Therefore, this research will not only serve to draw a more detailed image of the situation in EEMCS generally and HMI specifically but also be used to check the conclusions drawn from the literature against the situation at the University of Twente. At the end of this chapter, the first research question is answered and design implications for the ethics assessment tool are drawn which point towards an adaptable questionnaire that supplies users with information and examples relevant to them.

4.1 Modelling the ethics assessment process

The aim of the research towards practices around ethics assessment in and around HMI is to supply a more detailed image on several topics and check the conclusions drawn from the literature. To make sure all parts of the process are addressed a model of the ethics assessment process was made to serve as an overview. The model shows the current understanding of the ethics assessment procedure in HMI and its stakeholders (see Figure 1).

In this model, actions are put in normal squares, and stakeholders are marked by bold squares. Inputs and outputs of information and/or materials necessary for the actions are shown with arrows. The white area in the center marks the part of the ethics assessment process that this master thesis focuses on. The model of the ethics assessment process forms the basis for the research towards practices around ethics assessment in and around HMI.

Furthermore, each step can also be approached from different perspectives, because multiple factors can influence a person's results when trying to complete an ethics assessment. It is not only important that they are capable of successfully doing a certain step, but also if they think the step adds something of value, or what kind of help they want to have or seem to need at that point. The different perspectives on the ethics assessment process that are seen as valuable to address are elaborated on below.

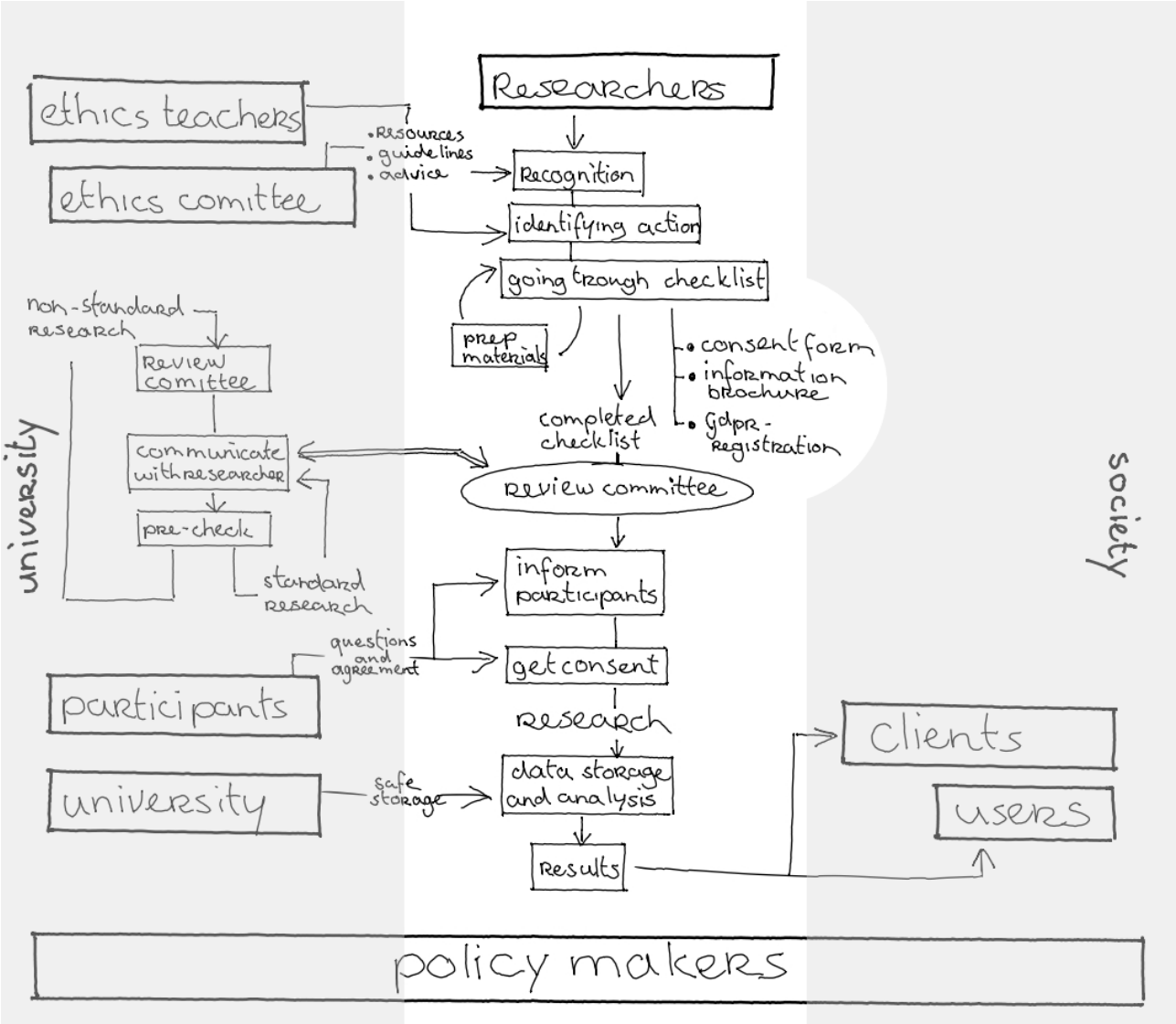


Figure 1: Model of ethics assessment procedure

Problems that occur during the HMI ethics assessment process. The problems that occur in the field of HCI in general probably also occur in the HMI department. However, it is also possible that other problems are experienced, or known problems are experienced differently due to differences in working culture, capabilities of researchers, ethics committee, or other circumstances. Next to that, comparing real-life experiences with the insights from literature can highlight possible holes in the literature.

Types of assistance people use to solve these problems. To get a better picture of what kind of help people would like to have it is useful to see what kind of help they already find themselves during the ethics assessment process. This gives insight into the practices around the ethics assessment process and what sort of help people benefit most from.

Perceived efficacy in the ethics assessment process. According to Section 2.4.1, a low efficacy in the ethics assessment process affects the willingness to engage in it. Therefore, having an idea of the perceived efficacy of people regarding different parts of the ethics assessment process can help in finding out where more assistance or assurance is effective.

Peoples actual abilities in going through the ethics assessment process. Knowing what part of the ethics assessment process people do poorly or very well in, also helps in determining when assistance is needed most and what kind of assistance is needed to improve the output of the ethics assessment procedure.

Perceived importance of the assessment process. If people see the ethics assessment process as an important part of their research, they are more likely to pay attention to it. This will result in a better outcome of the process. If the perceived importance of (a part of) the ethics assessment process is known to be low, advice or assistance might be adapted to highlight the importance.

4.2 Method

To collect the right information from people participating in the procedure, the interview method of semi-structured interview combines with a structured interview part, to gather participants' opinions on ethical dilemmas. This method was selected because the first part allows for some flexibility to collect people's opinions and attitudes towards the ethics assessment procedure, while the second part allows for gathering qualitative information about the ethical judgments participants make.

4.2.1 Participants

The ethics assessment process has many different types of stakeholders who can offer different perspectives on the topic. These stakeholders and some of their relations to the ethics assessment process can be seen in Figure 1. To keep the research toward practices around ethics assessment in HMI into the scope of this graduation project, only the stakeholders who engage in the ethics assessment process (semi)actively and at least have some experience with the process will be included as participants. This group of people is quite diverse, so they have been split up into different categories, which are described below.

Creative Technology students doing or just having finished their thesis

Students who are finishing their Bachelor of Creative Technology and starting their master are generally between 20 and 23 years old. They have completed (almost) all courses in their bachelor's, so their academic background is similar. They are working on either finalizing their bachelor thesis, or they just started with the first courses of their master's. During their bachelor's and their graduation assignment, almost all of them have been confronted with the ethical committee at least once, and they have had a course in ethics during the preparation of their graduation assignment. Therefore, they have some experience with going through the process of ethical assessment and thinking about ethics.

Interaction Technology students

Interaction technology students are generally between 21 and 24 years old, have mixed nationalities, and come from varied academic backgrounds. Some of them will have a Bachelor of Creative Technology while others might have a bachelor from a different university or a bachelor of applied sciences. Therefore, their abilities in completing the ethical assessment procedure are probably quite diverse. Depending on the courses they have already taken they may have had experience with doing ethical assessment and the ethics committee during their master's.

Ph.D. students, researchers, and supervisors

The group of Ph.D. students, researchers, and supervisors is very diverse in age and background. However, they all have one thing in common; they engage in ethics assessment both through their own research and through their role as teachers and supervisors. This enables them to provide an interesting perspective on the way students engage with ethics and the general view on ethics assessment within the department.

Members of ethics committee

Members of the ethics committee are most likely experienced researchers with either a background in ethics assessment or experience with it through being in the ethics committee. They will probably have extensive knowledge about guidelines, the opinion of other researchers towards ethics, and problems that might occur in the ethics assessment process.

4.2.2 Recruitment

Before the research started the research proposal was approved by the EEMCS ethics committee. After approval was received, participants were recruited in various ways. Students were asked to participate through WhatsApp and were sent the necessary forms and information once they showed interest in participating, while researchers and ethics committee members were asked via an e-mail that already included these. In total, 24 people were recruited of which eight older bachelor/master students, six ethics committee members, and 13 researchers. Note, that this counts up to more than 24, since three of the ethics board members were also interviewed as researchers.

4.2.3 Demographic

A total of 24 participants were recruited. Among the participants, there were eight bachelor/master students, six ethics committee members, and 13 researchers. Note, that this adds up to 27 instead of 24, since some ethics board members were also interviewed as researchers. This does not mean their opinion is counted twice, but that they also answered the question specific for researchers. The group researchers contained 8 Ph.D. students/Post-Docs and five (assistant) professors. No age was recorded from the participants, but the population of this participant group seemed relatively young compared to the general population of HMI researchers.

4.2.4 Design interview guide

As a basis for the interview guide, an overview of steps of the ethics assessment and their different perspectives (as stated in 4.1) was created in the form of a matrix, which can be found in Appendix A.

For each group, the steps of the ethics assessment procedure in which they are involved were selected (see Appendix A). Based on these, three different interview guides were designed, so the topics discussed with each participant would be relevant to them. The design of the interview guides was done based on good practices for making an interview guide described in literature (Newcomer, Hatry, & Wholey, 2015; Jamshed, 2014). A sample of the contents of the guide can be found in Appendix B. The first part of all three guides consists of core questions followed up with associated questions. These core questions could easily be extracted from the matrix since they already contained the main questions for each step. Associated questions are based on the different perspectives linked to these steps. For example, the core question for the step *"prepare necessary materials"* is: *"Are researchers able to prepare the necessary materials?"*. Examples of associated questions, in this case, are: *"Do you know what forms or materials are necessary for ethics approval?"* and *"who would you ask for help or advice?"*, which are linked to the perspectives of "knowledge/abilities" and "sources of help". additionally, to gain information on how people assess dilemmas in HMI, at the end of the interview for students and researchers five ethical dilemmas were added to have participants put them in one of three categories: "ethical", "questionable", or "unethical". After judging the dilemma, participants are allowed to comment on their judgments. The validity of the dilemmas in the HMI context was discussed with Dennis Reidsma, the daily supervisor of this thesis. For a last round of improvement, the interview guides were pilot tested with two EEMCS students that did not fall within any participant group. This led to some minor improvements in phrasing.

4.3 Procedure

Each session was conducted through a video conference due to the COVID-19 regulations. However, only audio was recorded. At the start of the conference, participants were reminded of the fact that only audio would be recorded and asked if they had any questions about the research. After all, questions were answered satisfactorily, the first recording was started to record informed consent as described in the consent form and information brochure approved by the ethics committee (reference number: RP 2021-28). The informed consent recordings were saved with the initials and last name of the participant. After this, a second recording was started to record the interview. All interviews were conducted according to the interview guide designed for the participant group(s) participants belonged to. At the end of the interview, participants were allowed to add their thoughts and comments if they wanted to. When the interview was over recordings were saved anonymously and separate from the consent recordings.

4.3.1 Analysis

Each audio recording was analyzed by summarizing it per question of the interview guide without any personal identifiers. The resulting summaries of each participant were combined per participant group, which resulted in one file per group with answers from every participant in that group. Then, for the groups "students" and "researchers", similarities and notable differences within the group were identified based on the participants' answers and added to the matrix in the best fitting "perspective" column. In case the researchers mentioned something about the behavior of students, it was also added to the matrix of the student group in a different color. This way students' own experiences and researchers' experiences with them in the role of supervisor are distinguishable. For the "ethics committee" group the first step was executed in the same way; however, the similarities, differences, and interesting viewpoints distilled from the answers were put in a text document instead of a matrix, since their answers and insights did not fit the matrix structure. Lastly, the data gathered from the ethical dilemmas were compiled in a table per group to be able to compare judgments between groups.

To find overarching themes for the discussion next to the more detailed results in the matrices, all results were read thoroughly. At places where participant groups discussed similar topics, their answers were compared to find agreements and contradictions. Findings from this part of the analysis are not included in the results but are used in the discussion.

4.4 Results

Interviews took generally between 30-40 minutes with some taking up to 55 minutes. Approximately 17 hours of audio was recorded, excluding the consent recordings. it is important to note that due to time constraints of the 3 ethics committee members who were also interviewed as researchers, only 1 was also asked to give their opinion on the dilemmas.

During the interviews, participants supplied a great quantity of useful information, suggestions, and interesting views. There is a vast amount of results that are too elaborate to put here. The interpretation of these results is discussed in the next section. Nevertheless, the full results are definitely worth the read if you want to know more about the state of ethics assessment in and around HMI and possible improvements. They can be found in Appendix C and D. Data on the comparison between the judgments of students and researchers is compact enough to be mentioned here (see Figure 1). The similarity of colors between students and researchers indicates agreement. For a full description of the dilemmas themselves see Appendix B. A shortened description is included in the notes on opinions below.

	Ethical		Questionable		Unethical	
Dilemma	students	researchers	students	researchers	students	researchers
1	0%	0%	30%	9%	70%	91%
2	14%	17%	28%	56%	58%	27%
3	0%	0%	0%	9%	100%	91%
4	58%	54%	28%	46%	14%	0%
5	85%	64%	15%	36%	0%	0%

Table 1: Assessment of the five dilemmas by students and researchers. Note that all percentages have been rounded to an integer number, and 11 participants are included in the "researchers" group instead of 13.

Notes on researchers' opinions on the dilemmas

Dilemma 1 - *Illiterate adults say it is not necessary to read the consent form aloud because they trust the researcher*: Participants were always quick to answer this with something like "no that's not okay because doing research without participants at least having the chance to hear the information is unethical" or "just read it to them and say that it is necessary for participation".

Dilemma 2 - *Participants wife supplies valuable insights for the experiment but is not an official participant*: When the dilemma was deemed "ethical" by the participant it was in most cases followed by including it in the research in a different way. such as anecdotal evidence. Most participants who answered "questionable" added that it would be possible under certain circumstances such as signing a consent form afterward.

Dilemma 3 - *Children forgot the signed consent form at home, the researcher still lets them participate*: Most thought this unethical, which is fully in accordance with regulations about this from the ethics committee. The only person saying questionable was thinking of the kid(s) feeling excluded. The general opinion was that if this happens your consent process is bad.

Dilemma 4 - *Same as dilemma 3 but no data is recorded of these children*: The participants who answered "questionable" were mostly concerned with the nature of the electronic toy and taking away time from learning without permission. Most think this to be "ethical" and see it as a good solution to not let these children feel excluded.

Dilemma 5 - *Research with prototype controller with handicapped adults who fully understand they may experience mild pain during the experiments and give informed consent*: Participants had a lot of good questions about this dilemma before they answered. They were considering what illness a participant had if a doctor had looked at it if everything was done to minimize pain, etc.

Notes on students' opinions on the dilemmas

Half of the students commented on the dilemmas. Others just gave their opinion and continued.

Dilemma 1 - *Illiterate adults say it is not necessary to read the consent form aloud because they trust the researcher*: Students who answered "unethical" agree that you have to read it to them so they have the chance to listen. If they do not want to hear it participants can choose to sit through it and not listen. That's their choice. Students who answered "questionable" bring up the same concern.

Dilemma 2 - *Participants wife supplies valuable insights in the experiment but is not an official participant*: Participants who answered "questionable" were looking for solutions to still be able to include it such as letting the woman sign a consent form. None of the students who gave one of the other options as answer commented on their reasons why.

Dilemma 3 - *Children forgot the signed consent form at home, the researcher still lets them participate*: All who commented already suggested the 4th dilemma as a solution.

Dilemma 4 - *Same as dilemma 3 but no data is recorded of these children*: Some who answered "questionable" were wondering if the electronic tool could hurt the children in some way.

Dilemma 5 - *Research with prototype controller with handicapped adults who fully understand they may experience mild pain during the experiments and give informed consent*: Participants seem hesitant to have an opinion about this. One mentions that this is the reason they like the fact that the ethics

committee is there to assist them.

4.5 Limitations and biases

There were a few limitations and biases were recorded during this research. First of all, for the research towards practices around ethics assessment in HMI, a second method was considered which entailed observation of participants going through the ethics assessment process. This should have yielded a more objective assessment of participants' skills. However, the interviews yielded sufficient information on the topic of participants' skills, and this second method was abandoned.

Secondly, the participant population of researchers was relatively young. This might have influenced results because younger researchers have been working with the ethics assessment procedure their entire careers. A reason for this could be that senior staff members do not conduct their own research anymore but supervise larger projects in which junior staff does the work with participants. However, it could also be that senior staff members of HMI have more responsibilities and therefore less time to participate in an interview.

Finally, participation in this study was voluntary, which may have biased the participant population to be more interested in the ethics assessment procedure than average. This is because people who are already interested in research ethics or think the ethics assessment procedure is important were probably more likely to join. This should be taken into account while interpreting the results.

4.6 Discussion

This section aims to interpret and discuss the data gathered during the interviews and compare it to the findings from the literature. From the data, it quickly becomes clear that the process of ethics assessment has room for improvement. This is especially true in the areas of clarity and ease of use for people with little to no experience with the ethics assessment process since these things are mentioned regularly in all participant groups. These points are specifically mentioned by people who come from outside of HMI, because their research generally has more indirect ethical issues which can be more challenging to address properly and because they generally have little to no experience with it. However, people are also clearly willing to improve this, since where there is a lack of clarity, many researchers have created their own examples of forms, etc. to make the process less confusing to go through for themselves and the students they supervise.

4.6.1 Quality of submitted requests

It was reported by students, researchers, and ethics board members, that the quality of ethics assessment requests varies widely per individual. However, committee members also noted that researchers generally deliver better and more consistent quality than students. This matches the opinion of researchers acting as supervisors, on the quality of requests from master students and Ph.D. students. Supervisors also indicate that the quality of a student's work depends not only on their skills or experience but also on how important they deem the ethics assessment process and the forms they are working on. If they find part of the process important, they seem to be more aware of what actions to do to complete it.

4.6.2 Attitude towards ethics assessment process

Almost everybody participating in ethics assessment says they are interested in ethical implications of research and willing to work together on making improvements, which might be overestimated because of the biases present in the participant population. However, there is some frustration present in all groups. According to both researchers and students, this frustration seems partly caused by the

process but also by shortness of time and little knowledge about or experience with ethics assessment. The majority of students indicate that their, or their peers, bad experiences with the process cause them to delay going through the process. Although, it has to be noted that students' attitudes vary greatly per individual. Students also reported that the time it takes to receive approval from the committee takes too long because at the time they would have to go through the ethics assessment process to make sure their research is accepted on time, they are not entirely sure yet about the exact details of their research. This leads to either ethics assessment requests being phrased very broadly, changes to research without re-submission of the request, or requests that are submitted too late to receive approval on time.

Researchers did not report any negative effects on the ethics assessment process due to their frustration apart from a general dislike of the process. On the contrary, one of the ethics committee members noted that there is a small minority of researchers who seem to avoid the ethics assessment process. Since it is unclear if these researchers participated in the study, it is unclear if they were simply not interviewed, or if they did not want to talk about this fact during the interview. Also, two ethics committee members mentioned that at the University of Twente there sometimes seems to be some apprehension among researchers towards the ethics committee, because they have the feeling they might not be able to execute the research the way they wanted because of them, and that this might affect the quality of their results. This suggests that the trade-off between ethical research and results has indeed some effect on the attitude of researchers towards ethics assessment.

Lastly, ethics board members themselves generally indicated to be motivated to contribute to the ethics assessment process and to like doing related tasks. However, sometimes they also experience negative feelings towards it. If they are frustrated, feelings of under-appreciation and lack of time to complete all requests play the most important role. However, not all experience this frustration equally. One last interesting thing to mention about the ethics committee members is that most see the goal of their work to be not only about ensuring ethical research but also about teaching students and others about how to recognize and approach ethical issues.

Assessing real ethical dilemmas is taken seriously, especially by researchers. Except for some researchers from outside HMI, they generally seem to be very thorough in their deliberation about possible ethical issues and the factors that play a role in them, as can be concluded from the results shown in Table 1. From their comments on the dilemmas, it is clear that they are quite careful when assessing a situation on being ethical or not. According to the same figure, students in the Creative Technology Bachelor and Interaction Technology Master generally make the same judgments about these situations as the researchers, but are a bit less hesitant and take less time to think through them. This is supported by the way researchers and students assessed the same ethical dilemmas. This attitude also became clear when participants talked about how they deal with data storage and analysis. Researchers all answered that they adhere to the protocols in place for storage and valuation of data very closely, while the majority of students indicated they are a bit freer with this. However, a few of the student participants indicated they do adhere closely to the protocols if they are dealing with sensitive data. These were also the student participants who thought it was very important to keep personal data safe. Clearly, students put more effort into it if they think it is more important to do so.

4.6.3 Comparison with literature

When looking more specifically at the challenges mentioned in the literature, as described in section 2.2 it is clear that some but not all challenges seem to be present at the HMI faculties, and that some challenges do appear but in a slightly different form. Below, each one determined from the literature is compared to the findings from the interviews.

The guidelines can be difficult to understand and follow properly. This challenge is also present in the HMI faculty, especially for those with little experience as is clear from the answers from individ-

ual students and researchers. If they indicated to have little experience, they almost always indicated that they found some part of the process unclear or difficult. There are some examples of people not understanding questions and the explanation, consisting of 30 pages of text, is said to be difficult to start on.

The guidelines are written in a language that is hard to understand. This is not the case in the HMI faculty. When the explanations are read, which only a few participants said to have done, they are clear.

The guidelines are difficult to locate This is partly true for the EEMCS guidelines. They are not necessarily hard to find but according to some of the older students, there are two places where you can find guidelines that are not completely the same, which confuses people.

Standards and regulations are sometimes inconsistent and too general. This challenge is present in the HMI faculty. According to two of the ethics committee members, questions are phrased in such a way that they apply to all research, but this makes it difficult to answer them. They are too vague. This is supported by the fact that one researcher who is not originally from HMI said they struggle with the question of whether an ethics request is even necessary for their research.

There are few (if any) examples of how these principles are applied in practice. This challenge is present in the HMI faculty. Both students and researchers mention examples are missing and/or that they (would like to) use examples as a source of help. However, people are already trying to solve this shortcoming themselves. Most Ph.D./researchers have made their own examples and students google them or get one from the BMS faculty (University of twente, 2021).

Some of the guidelines mostly deal with usability testing and handling of videotapes. This does not seem to be a problem within the HMI faculty. There are many issues with guidelines but this is not one of them. However, the guidelines are mainly focused on working with participants. According to two ethics committee members, this can make it vaguer and/or confusing for students and researchers originating from other EEMCS departments or HMI students who do not work with participants, but have to address other ethical issues such as biases in their data science project.

Guidelines do not promote working with vulnerable target groups. In one way this is the case in the HMI faculty, because vulnerable people are not included in research not specifically focused on them according to both researchers and students. However, many researchers said they work on research specifically for vulnerable groups very often. So, they are not excluded from research as a whole but they are also not part of the "normal" population.

Researchers' skills with preparing materials are insufficient. During the interviews, participants' skills were not compared to some sort of standard, so it cannot be judged how (in)sufficient they are. However, skills vary a lot. Important factors in this, as indicated in 5.6.2, are prior experience, interest in the process, and perceived importance of it. Researchers on the fringes of HMI and in other EEMCS faculties seem to struggle more because ethics assessment is quite new to these groups. In HMI this is less the case, although some individuals are still clearly better equipped for ethics assessment than others. Most pick up skills in ethics assessment somewhere in their master's if they are at least mildly motivated to do so.

Review boards are sometimes contra-helpful This challenge does not seem to be present at the HMI (and the rest of the UT). According to both students and researchers, it can be hard to reach them over e-mail but, if you get a hold of them they are very helpful.

HCI is exceedingly situational while anticipatory processes are not adjusted to this. This is a problem present at HMI (and the rest of the UT) but it also gets resolved quite well already. For example, some students have mentioned difficulties with the fact that they have to know every detail in their procedure beforehand. According to the interviews with students, these issues are resolved, sometimes with advice from supervisors, by writing down something more general and making changes within the bounds of the ethics approval.

Existing procedures do not provide assistance in dealing with situational ethics after approval. This is a challenge present in HMI, although this does not seem to cause many problems since re-

searchers and students are quite capable of dealing with unexpected issues in a fitting manner. This is mostly due to good people skills and experience and not due to specific training since nobody indicated they were ever trained in this.

4.6.4 Specific point of struggle in the process

As for the first research question: at which specific points in the process of gaining ethics approval do researchers struggle and why? The results show not only general problems but also the specific steps in the process people struggle with most. This is different per group. To make the text in this section more readable, the specific points of struggle are highlighted in **Bold**.

In the student group, the step at which participants experience the most difficulty varies quite a lot. However, most participants experienced difficulties while **writing the consent form and information brochure**. Even if the expected content of these forms was clear they did not know where to start. Furthermore, supervisors mentioned that some students have problems with **recognizing the need for ethics assessment**. It is not that they cannot do it, they just forget about it. Also, **applying feedback** and proper **data storage** are problematic in a few cases. According to supervisors, the main reasons why things go wrong in the ethics assessment process are students thinking a step of ethics assessment is not important, or a lack of experience in going through the process.

However, researchers experience fewer issues during the ethics assessment process. Although, there is a clear difference between HMI researchers who have a lot of experience with human subject research and HMI researchers who have not. The second category struggles with the step of **identification of necessary actions** to complete a proposal and with **preparing materials**. This is however only a small minority of researchers. Most researchers seem to only experience issues at the step of receiving feedback. at least half of the participants thought it takes too long to **get feedback** an/or get a reply on the improved version of a proposal. This fits with the general issue mentioned that for low-risk, standard research such as expert interviews the procedure is too elaborate.

4.7 Implication for the ethics assessment tool

The possible usefulness of an assistance tool for the ethics assessment process to decrease frustrations and improve the quality of ethics assessment requests is grounded in both literature and research towards the practices around ethics assessment in HMI. Many of the challenges identified in the literature are confirmed to also be present in the HMI department. Furthermore, the specific points of struggle identified and the solutions for these struggles people already found themselves, provide further arguments for the possible helpfulness of such a tool. In short, designing an assistance tool for ethics assessment seems like a sensible next step in improving the quality of ethics assessment in HMI.

4.7.1 Design considerations for a balanced advice tool

The combination of literature and research towards the practices around ethics assessment in HMI also provides general design considerations for an assistance tool. These considerations provide a starting point for the design of the tool which will be addressed in detail in the next chapter. The different design considerations are highlighted in **bold**, to allow for a quick overview.

First of all, **there needs to be a good balance between helpfulness and sense of ownership of the completed ethics assessment request**, since the goal of going through the ethical procedure at the University of Twente is not only about ensuring research is conducted ethically, but also about learning how to do responsible research and deal with ethical issues, according to ethics committee members. If the only goal was an optimized quality of requests and minimizing effort on the part of

the researcher/student, it might be tempting to provide them with almost complete information and forms in which they only have to fill in some information. However, this does not prepare them for independent ethics assessment later in their career and does not prompt them to identify and reason through ethical issues themselves. But, letting them struggle on their own without assistance only leads to frustration and a negative view of the procedure. Therefore, the balance between assistance during the process and ownership of the result is very important for this tool. It should assist when necessary for going through the process smoothly, while still activating people to think through dilemmas themselves when possible. Later, it may be useful to let users decide themselves on the amount of helpfulness they would like, but this is something for future work.

Furthermore, **the ethics assessment tool should contain questions tailored to the proposed research methodology of the user**, so they can be assisted with dealing with the ethical implications specific to their choice of methodology. This can be concluded from the challenges from the literature and is supported by the research conducted in HMI, which indicates that guidelines are too general and hard to follow and that some questions in the ethics assessment preparations do not apply to all of the research that is conducted, confusing people going through the procedure.

Also, **it should give examples of how to structure certain forms or answer complicated questions**, since the lack of examples is addressed quite often during the research conducted in HMI and is also mentioned in the literature. Most students struggle to some degree with the contents of the consent form and information brochure and many researchers have created their own examples of these forms due to a lack of them. So for these forms, examples should be a part of the tool.

Additionally, **the tool should also be flexible enough that one can go through it quickly** and has to fit into the ethics assessment procedure in such a way that it is not mandatory to follow. This is because some researchers are very experienced in ethics assessment and sometimes struggle with the time the entire process requires. These people could benefit from the ethics assessment process in general running smoother, but not from having another task added to their list in the form of this tool.

Lastly, **when people do use the tool it should lead to a higher quality ethics assessment request** because this could relieve some of the workloads of the committee members. If the tool can already provide answers to simple questions about the procedure, committee members will have more time to dedicate to assisting with complex ethical issues.

5 Application of design implications in assistance tool

In this chapter, the second research question is answered by using the results and design considerations from the previous chapter to design an ethics assessment assistance tool for students and researchers in HMI. Currently, there already is a tool available in which questionnaires can be both constructed, filled in, and generate example forms. It was developed by Amy de Lange and Giorgos Hadjidemetriou and commissioned by Dennis Reidsma as part of the improvement of ethics assessment in HMI because there was already a notion that such a tool would be needed for this assignment (A. de Lange, 2021). Since this tool fits the design considerations stated in the previous chapter, it will be used to design the ethics assessment assistance questionnaire. Therefore, the existing tool will be described in the first part of this chapter, followed by a discussion of the structure of the contents of the tool and a description of the final versions(s) of the advice.

5.1 Details of the existing tool

The goal of the current tool is to supply an environment in which flexible questionnaires can be made that probe the user about the details of their proposed research. This information can then be used to supply the user with an example consent form and information brochure, and possibly other advice, after completing it. It consists of an editor environment in which questionnaires and advice fragments can be drafted and published and a user environment in which published questionnaires can be filled in and example forms can be downloaded. The technical details of both environments will be described separately below. All information about this tool was supplied by the developers of the tool (A. de Lange, 2021).

5.1.1 data structure

A questionnaire consists of several linked parts. First, there are the sections. These form the main structure of the questionnaire. Each section has a section name and can contain one or more frames. Frames contain the information displayed in one step of the questionnaire. They can contain one or more different questions. A visual representation of this structure can be seen in Figure 2. First, there is the `frame_name`. This must be unique as it functions as the frame identifier in the database. The text displayed in the frame originates from the input in the `question_text`, `explanation_top`, and `explanation_bottom` fields. `Explanation_top` is meant to contain short pieces of information useful before reading the questions since it is displayed above the `question_text` in the frame. The `explanation_bottom` is more suited for more elaborate explanation not everyone needs to read when seeing the frame. Below this, the question(s) are displayed. Questions also have a unique key and label. The label is used to label the data from the question in the database. The key is used for activation rules which will be addressed later. Next to this, each question has a `question_type` which dictates what kind of answers can be given and in what way. Optionally, a question can have a placeholder text which is displayed as an example answer if the question type allows.

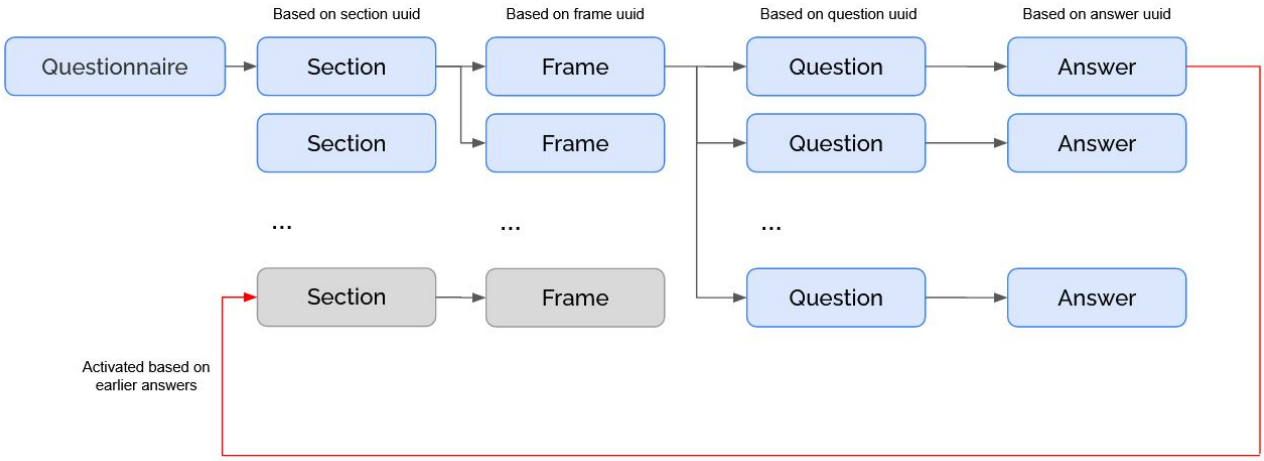


Figure 2: visual representation of data structure questionnaire

Furthermore, a frame can be displayed or hidden to the person going through the questionnaire based on the activation rules of the frame. Activation rules work based on answers given to previous questions in previous frames. As input, it needs the key of the question it has to activate on and the specific input given in the previous question it needs to activate on. Activation rules are structured like conditional logic statements. For example, an activation rule could state: `Participant_age_group == <18`. This indicates that this frame will only be displayed if at the question called `Participant_age_group`, the answer <18 was selected. One frame can contain multiple activation rules which can be linked with AND or OR statements relating to the same or different questions. So, it is possible to activate a frame only if multiple questions are answered in a specific way, to activate it if one of three relevant options is selected for one question or if one of several specific answers is given at one of several relevant questions.

Next to questionnaires, the tool also contains adaptable advice files, which are linked to a specific questionnaire. Advice files can be structured as a draft consent form, or contain general advice, depending on the wishes of the editor. Also, one questionnaire can be linked to multiple advice forms. A file can consist of a mix of static text, activation rule dependent text, and activation rule dependent text input given as an answer to an open question in a questionnaire. Activation rules here work in the same way as in the questionnaire. Once an advice file is generated, the related questionnaire answers are checked and text is displayed in a text file depending on the activation rules.

5.1.2 Editor environment

A user with editor rights can access the editor environment. Once inside this environment admin users can edit, create and publish questionnaires, advice files, and example forms/advice fragments; the workflow and navigation of the environment are pretty straightforward (see Figure 4).

If the "questionnaires" tab is selected, the "start screen" shows all published and draft questionnaires the user has are displayed. Here, they can choose to edit a draft to continue working on, copy an existing questionnaire to make a different version of it, or create a new questionnaire from scratch. Once a questionnaire is created or selected for editing the user enters the questionnaire editor. A full overview of the Questionnaire editor can be found in Appendix E. Now, they can drag and drop sections and frames in the desired order, insert questions and explanations and link them to each other with activation rules.

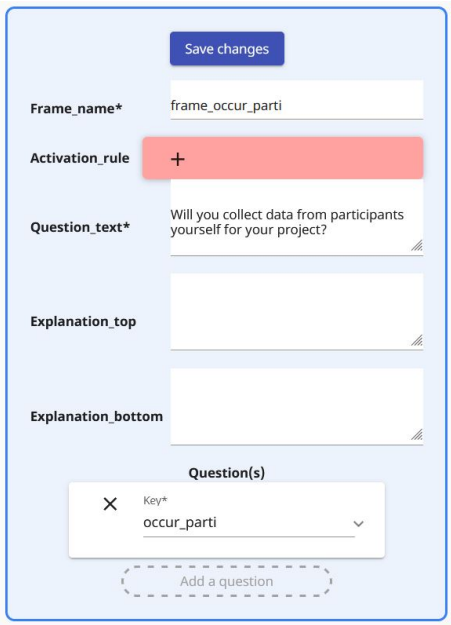


Figure 3: Example question

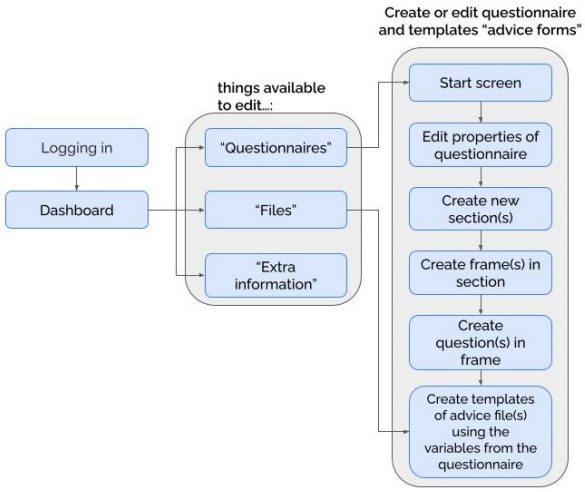


Figure 4: Simplified version of editor environment workflow in tool

After creating a questionnaire, example forms and advice files can be created as well under the “files” tab. Different activation rues can be added for different sections of the forms that toggle on one or multiple specific answers given in the questionnaire.

5.1.3 User environment

The workflow of the user environment is shown in Figure 5. Once a questionnaire is published, regular users can log in and fill in the questionnaire by navigating to "questionnaires" using the menu on the left. Here, they can select a questionnaire they have already been working on or start a new one. In a new questionnaire, the user starts at the first frame of the first section. An example of what a frame looks like to the user can be seen in Figure 6.

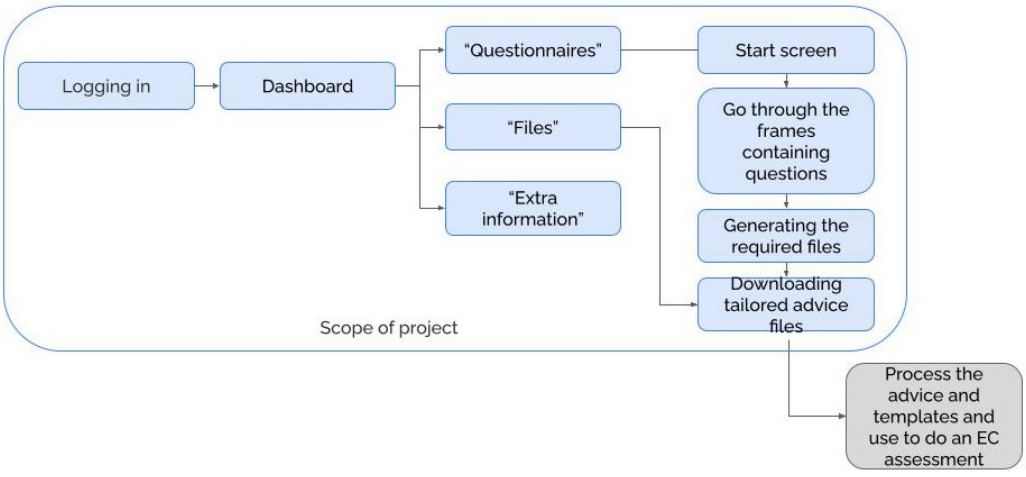


Figure 5: Simplified version of user environment workflow in tool

As you can see, some of the text of the `explanation_bottom` text box is not visible at first, but is only visible when the user wants extra information and presses the show more" button. Once the question(s) in a frame are answered the user can move on to the next frame in the order determined by the editor unless they select a specific section in the menu on the left. Depending on the activation rules

added, some frames and/or sections become invisible to the user, so they will not be confronted with unnecessary questions. If the questionnaire is complete, the advice files will be generated containing tailored advice based on the answers they gave in the questionnaire.

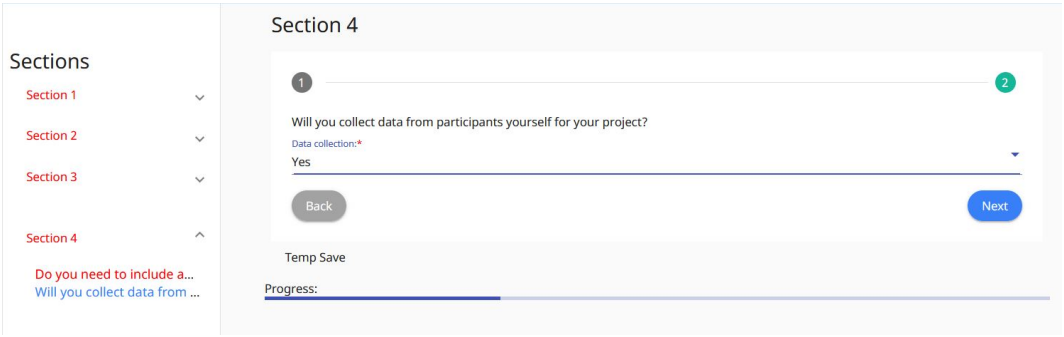


Figure 6: Example question

5.2 Shaping the questions and advice of the tool

The content of the tool was shaped according to the design implications from the previous chapter, to probe the user about the research they are planning to then provide fitting advice and suggestions.

First of all, the ethics assessment tool should contain questions tailored to the proposed research methodology of the user. To this end, the typical research patterns in our research group HMI were identified. For example, many of our students and researchers work on educational robots for kids with learning disabilities or autism. From these typical patterns, the most common research types were chosen, along with some ethically challenging options such as dealing with proxy consent from parents or working with kids who do not always clearly communicate their wants and needs, to be used for the prototype questionnaire. The full list of research types and the ones that were selected can be found in Appendix F. Based on these research types, questions were designed which, when answered by a researcher about their proposed research, can quickly rule out or confirm the presence of a certain research type.

The tool should also be flexible enough that one can go through it quickly, without answering unnecessary questions. To this end, the questions based on the selected research types were sorted based on what research types they are needed for and how many other questions they make redundant. With this information a flowchart was made with the more general questions first and more topic-specific questions later in the flowchart (Figure 7). For example, the question "Does your research include participants?" is put at the top because the answer "no" to that question indicates the redundancy of more than half of the other questions. By implementing this structure in the questionnaire, the user will be confronted with less inapplicable questions.

Furthermore, there needs to be a good balance between helpfulness and ownership of the completed ethics assessment request. Since it is not quite clear how much helpfulness is the right amount to assist the user as much as possible without impacting their sense of ownership too much, three different levels of advice were made. These are all implemented in different versions of the questionnaire, so it can be evaluated where this balance lies. Another design criterion is that the tool should give examples of how to structure certain forms or answer complicated questions.

This is also incorporated in the different versions of the advice. The **minimum helpful advice** version consists of short questions without any explanation and *point by point advice* about what topics to address in the ethics assessment and forms. This does provide users with an idea of what topics are important in their case, but does not provide any structure and only very minimum examples. The **medium helpful advice** version gives, next to the same short questions, some explanation about what

is meant by a question if this is unclear to the user. The advice consists of *a list of what topics to address with an explanation about why it is important* and what to pay attention to in the assessment and forms. Furthermore, a general structure of the consent form and information brochure is given, but the actual text has to be written by the user. The **most helpful advice** consists of questions with a lot of explanation and provides advice that *tells the user what to do in their case*. Much of the advice is incorporated in a draft consent form and information brochure, in which the user only has to add some necessary information.

However, the main purpose of the contents of the tool is not only to be helpful to researchers who have to complete an ethics assessment request, but also to improve the quality of the requests and educate researchers about good practices around ethics in HMI. Therefore, not only the quality of the resulting ethics request are important, but also the engagement with the ethics and educational value through the questions and advice in the tool. This was also an important consideration in shaping the question and advice of the tool.

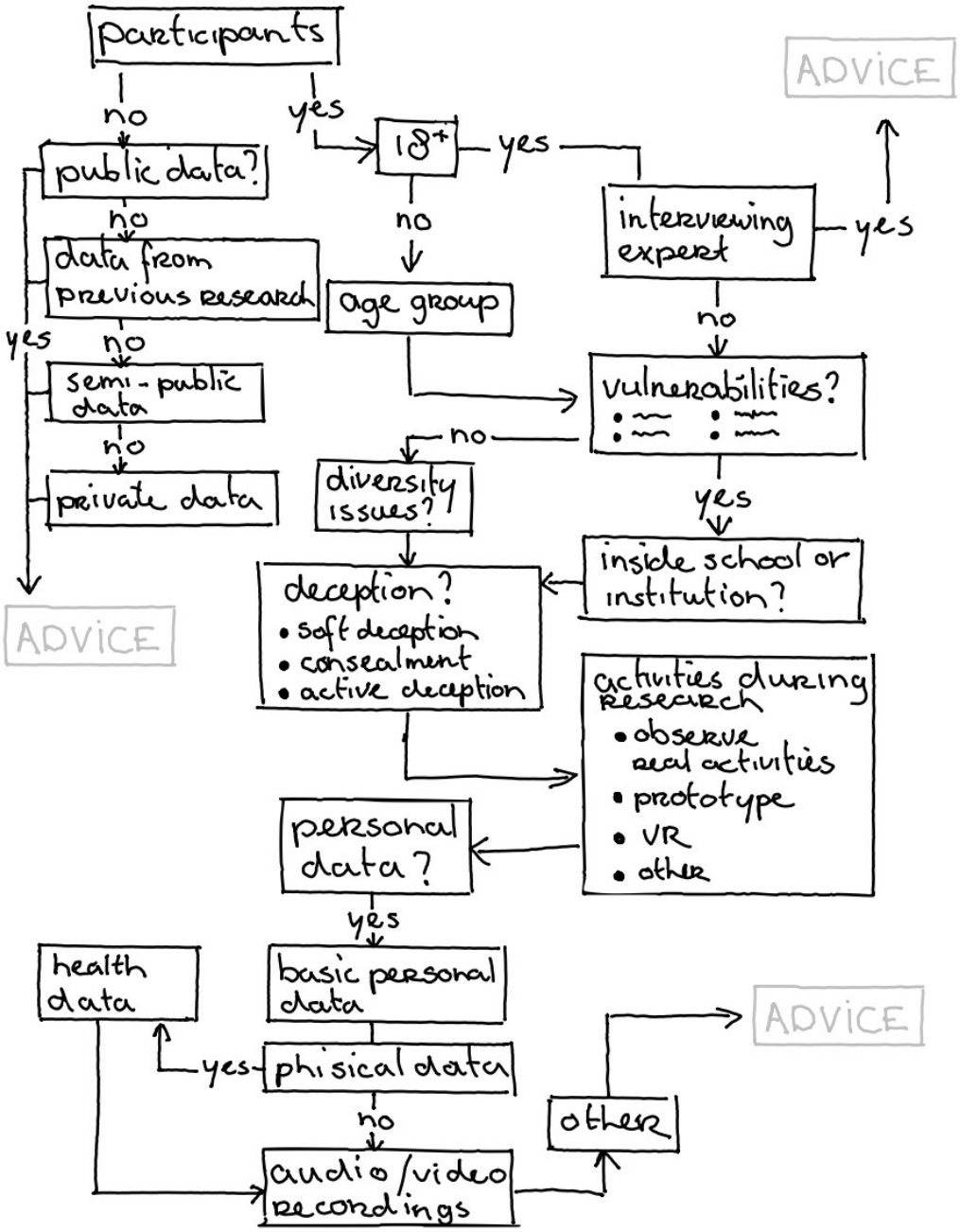


Figure 7: Structure of questions in the questionnaire. Paths in the flowchart are indicated with arrows and yes/no are conditional based on the answers in previous questions, other paths will be taken independent of previous answers.

5.3 Description of advice

As already mentioned in Section 5.2, because it is not clear yet what the ideal amount of advice is to balance helpfulness and ownership, three different types of advice were made according to the design considerations described in Section 4.7.1. Here these three levels of advice, as they were implemented in the tool, will be described in detail. The general structure of the questionnaires is relatively similar since all were based on the structure shown in Figure 7. First, users are presented with section 1: "participant demographic". Through the answers given in this section, research without participants, expert interviews, and research with participants can be identified. Based on this, users are forwarded to the questions relevant to them. Research without participants is immediately forwarded to section "Data gathering". Users planning to do expert interviews are done with the questionnaire at this point. Users planning on research with participants are questioned further about possible vulnerabilities and diversity of the participants, after which they are forwarded to section 2. Here they get asked about the research activities and setting. Everyone who finishes section 2 is also forwarded to section 3 in which personal data gathering is addressed. a the end of the questionnaire the answers can be submitted. After submission advice will be generated based on the answers given. Depending on the helpfulness level of the questionnaire, advice will be more or less elaborate.

5.3.1 Minimum advice

The minimum advice questionnaire is built up out of simple one-sentence questions which can be answered mostly with a yes/no choice or by selecting one or several multiple-choice options. When the desired option is not among the options of a multiple-choice question the user can select the other option. In this case, an open-ended question appears which can be used to describe their situation.

The advice provided when completing the least helpful questionnaire is also very minimal. It is only meant to remind the user of important aspects they have to mention in their ethics assessment and/or but does not explain how or why to do this. For example, minimum advice given when a researcher plans to use VR in their experiment could be: *"Some people can experience motion sickness while or after using VR. Take measures so that you are prepared in the event this happens."*

5.3.2 Medium advice

The medium advice questionnaire contains the same questions as the minimum advice questionnaire. However, short explanations of difficult terms are given below these questions, so users can easily get some extra help with understanding a question. This space is also sometimes used to clarify in what cases a question should be answered with yes or no. For example the question: "are you interviewing experts on their expertise?" is followed by a list of the conditions under which an interview becomes an expert interview.

The advice provided when completing the medium helpful questionnaire elaborates more on the reasoning behind the advice. The user is provided with an explanation about why certain topics need to be addressed in the ethics assessment and what aspects are important to address them properly. Furthermore, a description of how to write the consent form, information brochure, and other necessary documents is included in the advice. For example, medium advice given when a researcher plans to use VR in their experiment could be: *"In the questionnaire, you indicated that your research contains interaction with a VR environment. Some people can experience motion sickness while or after using VR. You have to warn them of this possibility in your information brochure and take measures so that you are prepared in the event this happens. You also have to make a stop protocol so you can stop the experiment in time if the participant becomes unwell. In the consent form and information brochure, this is addressed further. Also, a template for creating a stop protocol is provided"*

5.3.3 Maximum advice

The maximum advice questionnaire contains even more extensive explanations of the questions asked. Furthermore, some questions have been split into multiple questions for simplification. The most important difference between the medium and maximum advice is the advice provided once the questionnaire is finished. The advice is similar to the medium advice with added specific instructions about how to fill in the ethics assessment form. Furthermore, while the medium advice only provides suggestions of how to write the consent form and information brochure, the maximum advice provides a full template of these files in which only specific details like research goal, names, and duration have to be filled in by the user.

The advice from the maximum questionnaire could look like this: *"In the questionnaire, you indicated that your research contains interaction with a VR environment. This activity includes additional risks to participants. Some people can experience motion sickness while or after using VR. This must be mentioned in the information brochure and is included in the information brochure template at the bottom of this document. You also need to have a stop protocol in place, so you can stop the experiment in time if the participant becomes unwell. An example stop protocol is provided at the bottom of this document. Lastly, make sure you take measures to prepare for motion sickness when conducting your experiment such as having a bucket in the room and letting people sit down once they are done with the session"*

6 Evaluation

The goal of the evaluation is to compare the different versions of the advice and answer the third research question with an indication of how effective the advice tool itself might be. Since the advice is supposed to balance helpfulness to the researcher and the amount of learning and ownership of the researcher, these aspects will be the main focus during the evaluation.

6.1 Method

To compare the different versions of the advice, a method was selected in which participants prepare materials for a fictional ethics assessment. First, they receive a case description containing an example of typical HMI research. The case description can be found in Appendix G. During the session, participants pretend that they want to execute this research and need to prepare for its ethics assessment. After reading the case description, participants get the chance to ask questions if they have any. Then, they are asked to fill in the adaptive questionnaire of one of the advice versions: minimum, medium, or maximum, in the tool. At the end of the advice questionnaire, they can download advice files. The contents of these files are dependent on both the version of the questionnaire and the answers given. With the advice, the participants then have to fill in a simplified version of the SHORT FORM self-assessment questionnaire which is normally used to prepare for ethics assessment in HMI.

This method was chosen to mimic the real-life situation in which the tool is used, while still maintaining enough consistency in the task to compare between the different versions of the advice.

6.1.1 Experimental design

To compare the different versions of advice, the experiment is between subjects in which the only variable is the version of the questionnaire participants are using during their session. Participants themselves are not aware of which version they receive. Since there are three versions of advice, the participants will be divided randomly into three equal groups. The metrics are all recorded in post-testing, except for the demographics and the materials participants prepare during the session.

6.1.2 Participants

Participants in the evaluation were recruited from the group of first- and second-year create students because they are all possible future users of the tool and will all have almost no experience with the ethics assessment procedure. This minimizes the chance of possible biases due to the knowledge participants already have about the ethics assessment procedure and allows for data gathering on a possible user group not yet included in previous research for this thesis. To check if participants had little prior knowledge about the ethics assessment procedure, this topic was addressed in the demographic questions (see Appendix G).

6.1.3 Recruitment

Participants were contacted through WhatsApp, email, and by talking to them personally. If they showed interest in participation they were asked to fill in a form with their name and e-mail address. Then, they received additional information such as the information brochure, the consent form, and a link to select a time slot via this email address.

6.1.4 Procedure

The procedure is designed to mimic a real-life situation of the interaction with the self-assessment questionnaire. Each session starts by asking the participant for informed consent through a consent

form and information brochure approved by the ethics committee (reference number: RP 2021-204). Then, the ethics assessment procedure is shortly introduced and a participant orally answers a few demographic questions, which takes approximately 5 minutes. Then, the tasks the participant should perform during the session are explained. In short, the task consists of filling in the advice questionnaire in the tool according to the case description, downloading the advice generated by the tool, and filling in the simplified self-assessment questionnaire. A complete description can be found in section 6.1. The tasks take up to 40 minutes and participants are asked to stop and save their progress after this time. Then, they are asked to fill in an exit questionnaire via which they self-report their perception and performance of the different tasks, through reporting on their perceived effectiveness, sense of self-efficacy, motivation, and sense of ownership over the results are addressed. This takes around 5 minutes. Lastly, they are interviewed to dive a little deeper into their perception of the advice they received through a few oral questions which take a maximum of 10 minutes. All questions asked during a session can be found in Appendix G.

6.1.5 Measures and metrics

To compare the different versions of the advice and give an indication of how effective the advice tool itself might be, six different metrics were selected to track four different measures. Measures were selected based on their link to issues detected in the literature research and research in and around the HMI faculty. In this section, the reasoning behind the selection of these measures and the way these are tracked are elaborated. For clarity, the four measures are all highlighted in **bold** when they are first mentioned in this section, and metrics are highlighted in *italics*.

First of all, the **effectiveness** of the advice is evaluated, because one of the major points from both the literature and the research in and around HMI was that the quality of the ethics assessment is often lacking. The effectiveness of the advice is measured by both checking the objective *quality of the materials* prepared by the participants and the *self-reported effectiveness* dimension of the validated questionnaire in the user experience questionnaire handbook (Schrepp, 2015). The quality of the materials is determined by the number of questions answered correctly. The correctness of these answers is determined by comparing the participant's answers to a correctly filled-in ethics assessment form.

Next to that, since many students indicated in the interviews conducted that they did not feel motivated to start with the ethics assessment, **motivation** is tracked. The user experience questionnaire handbook's scale for *self-assessed motivation* is selected to track this. If a version of the tool seems to improve motivation, it is a good indicator that users might be more willing to engage with the ethics assessment process.

Furthermore, **self-efficacy** is tracked because users recognizing and dealing with ethical issues independently is one of the goals of the ethics assessment committee according to the interviews conducted. Next to that, self-efficacy also indicates how well a participant performs, assuming participants can reflect properly on their own performance during the assignment. To track self-efficacy inspiration was taken from a publication about surveying self-efficacy of teachers (Tschannen-Moran & Hoy, 2001). From this paper, it was gathered that questions about a specific topic or example are preferable when trying to track self-efficacy. This idea is reflected in the *perceived-competence scale from the C.S.T.D*, which was selected to use for assessing self-efficacy (CSDT, 2022b).

Lastly, **ownership** was chosen to track a participant's ability to deal with ethical issues independently, because it came up in the literature and interviews as a concept related to how much a person has contributed and/or feels responsible for a resulting ethics assessment. The definition of ownership that applies best in this context is ownership as a sense of control and self-determination (Furby, 1978; Pierce, Kostova, & Dirks, 2001). Therefore, it was decided to assess ownership via these two concepts. The questions addressing a *self-assessed dependability* from the user experience questionnaire

handbook were used to assess the control part of the ownership perspective (Schrepp, 2015). The self-determination part of control is tracked with the part of the *PCASS perceived choice scale from C.S.T.D* (CSDT, 2022a).

All metrics, except the quality of the material, are tracked through a questionnaire. All questions are answered by rating them on a scale from 1(totally disagree) to 7(totally agree), as suggested by both sources of questions, to lower the moderacy response bias. Questions were sometimes edited slightly to make it clear that the question was about the task participants just performed, instead of about life in general. The questionnaire containing these questions can be found in Appendix G.

6.1.6 Data analysis method

The non-parametric data gathered in the questionnaire and the amount of correctly answered questions are analyzed using the Kruskal-Wallis test, which is selected because it is suitable for non-parametric data with small sample sizes and multiple groups. The Kruskal-Wallis test outputs a significance level which indicates how likely it is that the data is not equally distributed.

The null hypothesis for all measures is: **With a confidence of 95% there is no statistical difference between all groups (P > 0,05).**
The alternative hypothesis of all measures is: **With a confidence of 95% there is a statistical difference between groups (P < 0,05).**

This test is normally followed up with a post hoc Dunn's comparison test to see which groups are significantly different from each other.

However, the data from the questionnaire does not fit the test requirements. To prepare the data to fit the requirements of the test, the data from the questions based on the user experience questionnaire handbook has to be transformed. First, the negative questions are inverted. See Appendix G for which questions are negative. Then, the answers of each participant are checked for randomly answered questionnaire answers, according to the rule of thumb suggested by the user experience handbook (Schrepp, 2015): If the distance between best and worst response between questions from this source tracking the same measure exceeds 3, a response is considered suspicious. If this is the case in more than one measure in a single response, the response will be deleted. Next, the internal consistency of all questions targeting one measure is evaluated using Cronbach's Alpha. If this turns out to be lower than 0.7 for a certain question, this question is generally not considered reliable (Schrepp, 2015; CSDT, 2022b). If there are less than three questions per measure with an alpha above 0.7, the three highest-scoring questions are used. However, results are less strong in that case. Lastly, an average score per participant for each measure is calculated by averaging the score of all questions tracking a certain concept. These scores are used as input for the Kruskal-Wallis test.

The qualitative data gathered in the interview and materials participants prepared during the study are analyzed by summarising answers and comparing the answers of the participants with each other to see if there are any common themes or other interesting comments.

6.2 Results

The total participant group (N = 26) was divided into the minimum advice group (n = 9), medium advice group (n = 9) and maximum advice group (n = 8). The study went generally smoothly apart from a few hiccups with the way the tool skips through the questions. This led to some confusion among participants, which was resolved by having them determine the order of questions manually. However, one participant in the minimum advice group was so confused by the tool and how it worked that they did not even get to downloading and working with the advice. Therefore, data generated by this participant was excluded from the statistical analysis. This led to one less participant in the

minimum advice group (n = 8). Most participants were not able to finish the shortened version of the checklist within the time given. However, since most did read and comment on the consent form and information brochure part of the advice, this is not considered problematic for the results. Checking of questionnaire answers of each participant for random entries as described in the method resulted in 3 suspicious entries. However, since for all three this was only the case for one measure, it was decided to keep all three.

Based on the reliability analysis as described in the method, three questions were deleted. Two matching questions were deleted from the dependability metric: *"The advice and other documents provided by the tool were... safe/unsafe"*. The third question deleted belonged to the perceived choice measure: *"I feel like I added value to the preparation for the ethics assessment"*. Table 2 reports the results of the reliability analysis before and after deletion. As you can see, after deletion of the three questions, all measures have $\alpha > 0.8$, except for dependability, which means reliability is good. Reliability of dependability is acceptable.

Metric	α **before**	α **after**
Motivation	0.89	0.89
Dependability (ownership)	0.76	0.88
Effectiveness	0.90	0.90
Self-comptence	0.83	0.83
Perceived choice (ownership)	0.70	0.76

Table 2: Internal consistency of all measures. Note that α is rounded to one decimal.

The average answer (1-7) per measure per participant can be found in Table 4 in Appendix H, which also includes the number of questions answered correctly in the ethics assessment form and the participant groups participants belonged to. Note that throughout the results section, numbers are used to denote the groups: 1 for maximum, 2 for medium, and 3 for minimal. Qualitative results of the interviews are reported in the following section.

Results of the Kruskal-Wallis test can be found in Table 3. For all metrics with $p < 0.5$ the null-hypothesis can be rejected, and the alternative hypothesis is accepted: **With a confidence of 95% there is a statistical difference for the measures dependability, self-competence, and perceived choice between groups (P < 0,05)**.

The distribution of correct questions and self-competence are also worth mentioning. From the distribution in Figure 8, it is clear that the number of correct questions is wide-ranging within all groups. The distribution of self-competence in Figure 10 shows the similar range of group 1 (maximum advice) and group 2 (medium advice), emphasizing the similarities between those groups. This is also the case for perceived choice (Figure 11) and dependability (Figure 9). In Appendix H, the two box plots of the remaining metrics can be found.

Metric	Group	N	Mean Rank	Kruskal-Wallis (P)	Dunn's Post-hoc**	
Motivation	1	8	14,6	0,129		
	2	9	15,3			
	3	8	8,7			
	Total	25				
Dependability	1	8	17,4	0,012	Groups	Adj. Sig. (P)*
	2	9	14,6		3 - 1	0,012
	3	8	6,9		3 - 2	0,092
	Total	25			2 - 1	1,000
Effectiveness	1	8	13,9	0,426		
	2	9	14,7			
	3	8	10,2			
	Total	25				
Self-competence	1	8	13,9	0,025	Groups	Adj. Sig. (P)*
	2	9	17,1		3 - 1	0,236
	3	8	7,5		3 - 2	0,022
	Total	25			2 - 1	1,000
Perceived choice	1	8	15,3	0,048	Groups	Adj. Sig. (P)*
	2	9	15,6		3 - 1	0,115
	3	8	7,7		3 - 2	0,080
	Total	25			2 - 1	1,000
correct questions	1	8	14,7	0,066		
	2	9	15,8			
	3	8	8,1			
	Total	25				

Table 3: Results Kruskal-Wallis and Dunn's comparison test. *Significance is adjusted by the Bonferroni correction. **Dunn's post hoc test in only performed on those metrics for which P < 0.05. Note that 1 = maximum, 2 = medium and 3 = minimal.

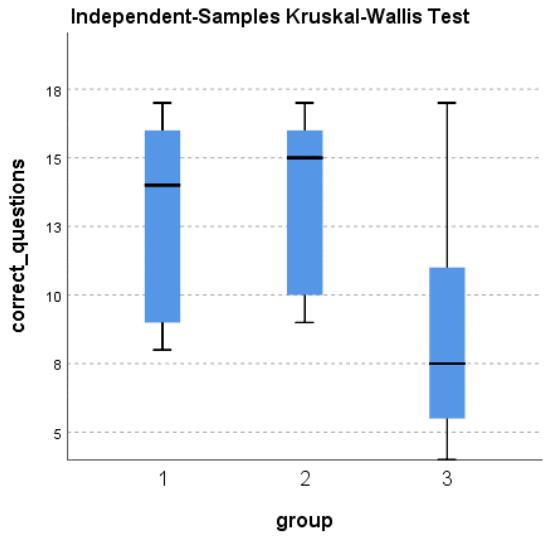


Figure 8: Boxplot correct questions

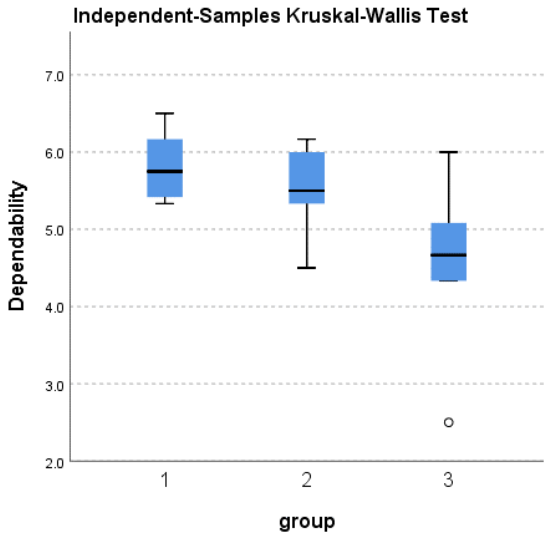


Figure 9: Boxplot dependability

Figure 10: Boxplot self-competence

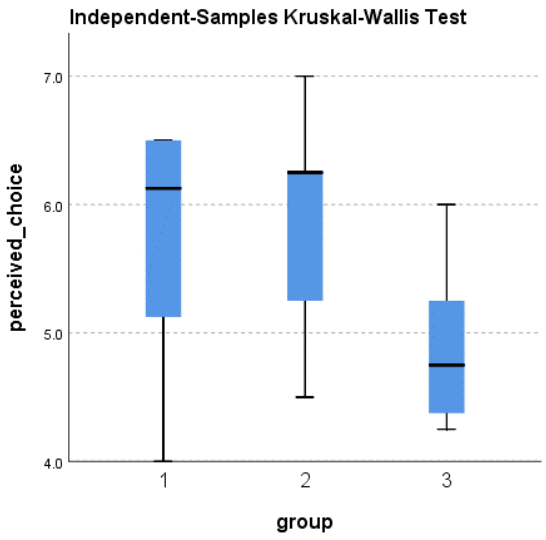


Figure 11: Boxplot perceived choice

When looking at the results of Dunn's test which was only conducted for the metrics with a significant statistical difference between groups (Table 3) to see where the significant differences between groups are, it is clear that there is no difference between the medium advice group (2) and maximum advice group (1) for the metrics dependability, self-competence and perceived choice (P = 1,000). Only for dependability, the most significant difference is between the maximum (1) and minimum (3) advice group. For self-competence and perceived choice, the difference between the minimum (3) and medium (2) advice groups is the most significant.

This, combined with the P-value of dependability of the comparison between groups 3 - 1 shows that: **with a confidence of at least 95%, dependability is higher in the maximum advice group than in the minimum advice group (P < 0,05)**.

Additionally, the P-value of the comparison between groups 3 - 2 for self-competence shows that: **with a confidence of at least 95%, self-competence is higher in the medium advice group than in the minimum advice group (P < 0,05)**.

lastly, the distribution of the number of correct questions is, when looking at the ranks in the table, likely skewed towards better performance with the medium and maximum advice. However, this result has to be interpreted cautiously, since the level of certainty that this measure is not equally distributed, is smaller than the desired 95% (see Table 3).

6.2.1 Qualitative results

Results from demographic questions

13 Participants were first-year Creative Technology students, 12 were second-year Creative Technology students, one was a third-year student following only second-year courses. Three students showed special interest in ethics and/or the ethics procedure, but most joined the study because they wanted to help out, or were curious how such a study works. Around half of the participants had a bit of experience with conducting research with participants. This experience mostly consisted of sending out a questionnaire in high school, or play-testing during the start of their second year. Only one participant had any experience with the ethics assessment procedure because he looked at the form in module 3 when their project group wanted to do some user testing. However, this idea was

not followed through because of time constraints.

Results from the materials prepared by participants

The self-assessment checklists participants worked on and the (parts of) consent forms some participants prepared during the experiment, yielded some qualitative data. First of all, there is a big difference visible between participants in the quality of the answers given in self-assessment form, even when they belong to the same participant group. These differences can both be seen in the number of correct answers and level of detail in answers, but also in how some participants seemed to struggle during the assignment. There was also one participant who struggled so much that his data could not be used in the evaluation. Furthermore, some questions in the self-assessment questionnaire seemed to be especially confusing for participants to answer correctly. question VIII. "Privacy, GDPR, and the possible need for DPIA" was only answered correctly twice by participants. Also, the question about the types of consent was answered incorrectly 1/2 of the time and the questions about deception were mentioned as very confusing by five participants. Lastly, preparing a consent form was very challenging for the few participants who got to that point. Even if they got some guidance in the medium advice it took them a very long time to translate that into some kind of form. They especially struggled with the type of language generally used in a consent form.

Exit interview

The exit interview consisted of four main questions. The first question was: "What do you think about the way the advice was given?" From the answers participants gave, it was clear all participants thought the provided advice was at least somewhat useful, no matter which participant group they belonged to. However, participants from the maximum and medium advice group were generally more positive. Good examples of this are these quotes from a **minimum** advice group participant: *"It was relatively difficult to estimate how detailed you needed to be. the advice was helpful, although some more elaborate advice would have been nice"*, **maximum** advice group participant: *"[Using the advice felt] as if you get a cheat sheet for an exam, it is great. I can imagine that doing this without it [the advice], it [filling in the self-assessment form] is difficult."* and **medium** advice group participant: *It is easy to make a story [answer to the question] out of the advice. The explanations of why you have to do something, [makes it] interesting and logical."*

The first question was followed up with: "What type of advice do you think is most useful to you? you can choose from: explanation of difficult terms, instructions or template on how to answer, applicable examples, a list of points to include and/or explanation about why certain topics were important." Most participants found it difficult to give only one answer to this question and mentioned two types of advice that they seemed to find equally useful, so the following percentages count up to more than 100%. Of the five different types, participants thought they would most appreciate receiving applicable examples (60%), followed by, instructions or templates on how to answer (32%), explanation of difficult terms (24%), a list of points to include (12%) and explanation about why certain topics were important (8%). Reasons participants gave for wanting applicable examples where: *"It gives a good idea of what is expected of you"* (**medium** advice group) and *"[it clarifies] what kind of language to use and how detailed you need to be"* (**maximum** advice group). Participants who thought templates and instructions were most useful, seemed to be mainly referring to templates for the consent form and information brochure: *"[Advice was] relatively clear, explanations of terms was very useful, but I would rather have a real template for the forms [consent form and information brochure]"*. When the explanation of terms was mentioned as valuable advice, participants mainly referred to the explanations in the questionnaire of the advice tool, which helps you to answer those questions more accurately.

The third question, "Was there anything unclear to you?" provided many specific parts which were unclear to a certain participant, but there is not one part that jumps out. However, Minimal advice

group participants almost always answered with unclarities about the self-assessment questionnaire, while participants from the medium and maximum group also answered with unclarities about the tool or the advice provided. General unclarities mentioned are: a lack of structuring per question in the self-assessment questionnaire and a lack of attention if the advice on the sub-questions in the self-assessment form. The follow-up question "if yes, what and how could it be improved?" was answered cautiously, or not at all. Some participants felt like they did not have enough experience to give recommendations, as is clear from answers like: *"I don't know how to answer that", "not really I think"* or *"I think maybe the part about research with children could be explained better?"*. There were a few exceptions in which the participants provided well-considered improvements on the advice or tool. Possible improvements of the tool mentioned were: *"...improve [functionality] of the show more feature", "...fix the way the questionnaire skips trough questions"* (a familiar problem) and *"...add some more elaborate explanations to some unclear questions"*. Since the last improvement was suggested by a participant working with the minimum advice which contained few explanations, it makes sense that they would suggest improving the questionnaire by adding more explanations. The question "What did you miss in the advice given?" did not yield any additional information, since participants answered it with the same answers and suggestions given at the previous question, or simply answered: *"I don't know"*

Lastly, the question "What part did you think was very useful?" yielded diverse answers, but every participant mentioned something that helped them to complete the self-assessment form. Only the advice on vulnerability (mentioned by 36%) was mentioned more often than other parts of the advice in answers such as: *"Vulnerable participants explains this well, I did not think about children as vulnerable participants".* (**medium** advice group) and *"The more elaborate explanation at the subject of different participants and activities were very useful."*(**medium** advice group). It is important to note that all participants who mentioned this, belonged to the medium or maximum participant group. Also, participants that received the medium or maximum advice mentioned quite often that they thought the examples provided in the advice and templates of forms were exceptionally useful for them. This fits with the answers given in the first question.

7 Discussion and conclusion

In this section, the implications of the results regarding the effectiveness and helpfulness of the advice tool are discussed. Based on this discussion, limitations of this research are addressed and suggestions for future work are made.

7.1 Discussion of results

In general, the advice on filling in the self-assessment questionnaire is seen as helpful by participants but does not solve all challenges experienced during this task. According to the interviews, a majority of participants thought the advice helped them when filling in the checklist, no matter which kind of advice they received. However, many had still significant trouble with some parts of the self-assessment questionnaire, as became clear from the materials they prepared. Also, the majority of participants thought that getting examples applicable to the type of research they want to conduct would be most helpful for filling in the self-assessment questionnaire and preparing other necessary documents. When asked why, reasons given by participants mostly came down to that advice gives an indication of what is expected, both in regards to the subjects that have to be addressed and the expected specificity of their answers. This is in accordance with the results form the research towards practices around ethics assessment in HMI, during which students indicated that they would like to have more examples while preparing for the ethics assessment.

However, skill level and needs seem to vary quite a lot given the large variations in the number of correct questions within groups (Figure 9) and qualitative results. During the sessions, there were a few participants who had no trouble at all with understanding the questions and coming up with satisfying answers, no matter which version of advice they got. But, there was also one participant who struggled so much that his data could not be used in the evaluation. This affirms the claim made in the research towards practices around ethics assessment in HMI, that the skills of students in completing the ethics assessment and therefore needs regarding advice vary greatly.

Furthermore, the mean ranks and results of Dunn's test (Table 3 of the self-competence measure suggest minimum advice promotes the least self-competence, and the medium advice promotes the most self-competence. However, it is unclear what specific feature of the advice caused this effect. However, since the qualitative results suggest that applicable examples are most helpful according to participants, and only medium and maximum advice contained these, the presence of applicable examples seems at least partly responsible for the increase in self-competence. The minimum advice does not assure participants that they were filling in the checklist the correct way, according to the low self-competence participants working with this advice reported. The qualitative results support this since some participants using the minimum advice did not find it elaborate enough.

Moreover, the sense of ownership does not seem to decrease when the amount of guidance in the advice increases. Both dependability and perceived choice, which are the metrics that give an indication about the measure ownership, do not decrease for the maximum advice, in relation to the medium advice (Table 3, Figures 9 and 11). It was expected however, that the sense of ownership would be lower in the maximum participant group since the maximum advice was thought to be too guiding. This does necessarily mean that this expected decrease in ownership does not exist. It could be that the maximum advice was probably not guiding enough yet to cause a decreased sense of ownership because first and second-year students struggled more than expected when filling in the self-assessment questionnaire. Maybe, advice that was even more guiding than the maximum advice could have shown the expected drop in ownership. The low sense of ownership in the minimum advice group could be explained by the fact that it does not seem logical to feel a sense of ownership over the answer if you do not understand the question. This theory is supported by the qualitative results. Minimum advice participants generally mentioned unclarities regarding the self-assessment questionnaire

itself, instead of unclarities in the advice. However, since ownership is a difficult concept to track, the possibility that the two selected metrics do not track the concept of ownership properly should also be considered.

Lastly, there seems to be no significant difference between groups in effectiveness metrics, but it is still suspected that this difference exists, since many participants working with the medium and maximum advice indicated the advice helped with properly answering the questions in the self-assessment questionnaire, and the majority of participants from the minimum advice group told they would have liked to get more advice on how to complete it properly.

All things considered, the results of the evaluation show that an advice tool such as the one proposed in this report can improve the self-competence of users in going through the ethics assessment procedure. Furthermore, there are signs that it could improve the quality of the materials prepared for ethics assessment. For the participant group in this evaluation, the advice itself is more effective when it is presented in the form of examples per question that show what kind of aspects are important for that question, and clarify what information a consent form or information brochure should contain.

7.2 Limitations and future work

The limitations specific to the first interview study were already discussed in Section 4.5. In this section, the general limitations of this thesis and limitations specific to the evaluation study are discussed.

The evaluation study had its expected limitations. First of all, during the evaluation not everything went as expected. The idea was at first to do pre-and post-testing for some of the metrics, but this idea was changed to only post-testing after the first few sessions with participants due to the discovery that the pre-test did not track the metrics intended. Furthermore, a study with a larger group of participants could have provided more conclusive results. Also, the tool changed the order of the questions during use. Fixing this could make interaction with it less confusing. Moreover, since no participant managed to complete the whole assignment given during the evaluation within the time provided, the time needed to complete the assignment was estimated too short. Apart from these, there are some more insightful limitations and future improvements to discuss.

Practical improvements of the tool

Problems encountered during the evaluation indicate possible practical improvements of the tool; which could all be implemented when the next iteration of the tool will be developed. First of all, making advice documents should be integrated into the tool. Now, all advice has to be put into one big word document and activation rules have to be written manually. this is very prone to errors and there is no feedback for debugging in case the advice form cannot be generated. If the creation of advice forms is integrated into the tool, the activation rules could be made into selection boxes or drag and drop elements, which would greatly improve the user-friendliness of creating and editing advice forms. Furthermore, options in the form creator to manage the order in which questions are presented to the user should be added. Now, questions are presented to users in seemingly random order. A more logical presentation of questions would greatly improve the user experience. Also, the different options for creating questions such as multiple-choice questions should be fixed, so they work together properly with the activation rules. Now, activation rules can only be triggered by only one option selected option, which greatly limits the possibilities when creating adaptable advice. Finally, a manual should be provided in the admin environment on how to create questionnaires and advice, so inexperienced users are also able to use the admin environment properly.

Method used to measure helpfulness and ownership

The method used for measuring both helpfulness and ownership in the evaluation relied heavily on self-assessment by the participant, which requires the participant to engage in a lot of self-reflection; which can be effective but is also prone to biases caused by the self-reflection skills of the participant. Furthermore, the metrics used to track ownership were quite difficult to self-assess and, as already mentioned in the discussion about the ownership measure, possibly did not track this concept properly.

There was also little literature available about tracking ownership, which could indicate the wrong concept was selected for reflecting a sense of responsibility for the outcome of the preparations for the ethics assessment procedure in participants. According to literature from the field of psychology which was only discovered after the evaluation study was almost finished, agency could be a more insightful measure to track a sense of responsibility for the outcome of a process (Gallagher, 2012; J. W. Moore, 2016). In this literature, agency is defined as the experience that a person is the one who is causing or generating the action and that a sense of agency is important for feeling responsibility for actions (Gallagher, 2012; J. W. Moore, 2016). If the agency a person experiences dunging the preparations of the ethics assessment process could be tracked, it could provide another indication about the responsibility someone feels for the end result. However, the agency is also hard to track, because participants rarely remember the original thought that prompted a certain action afterward (J. W. Moore, 2016).

This leads us to future improvements to the method of tracking these measures. Possibly, during a future experiment the Think Aloud Method could be used to get more insight into both agency and ownership since this method can be used to track cognitive processes (Jääskeläinen, 2010). The Think Aloud Method could also prove useful to improve the advice and get more detailed information on the other measures. Now, evaluation was done through self-assessment by the participant, but tracking cognitive processes could help identify specific problem areas in the advice and give more detailed information about its effectiveness. Lastly, it might also remove some participant-specific biases that were present when using the self-assessment method.

Processing of the results

The way data was gathered and then processed might have led to slightly skewed results, because of a confusion between Likert scales and Semantic differential scales during the evaluation. The results of the evaluation of the metrics motivation, dependability, perceived choice, and effectiveness were recorded with a Semantic Differential scale; which was the scale originally used in the validated question from the user experience questionnaire handbook. During processing, however, the negative questions were transformed which transformed the data to a Likert scale; which could have impacted the validity of the results. To understand this problem correctly, we first need to understand the differences between Semantic Differentials and the Likert scale. A Likert scale measures *agreement* with a certain set of statements. A participant reads the statements and decides how much they agree with it on a scale from "strongly agree" to "strongly disagree" (Schibeci, 1982; Friborg, Martinussen, & Rosenvinge, 2006). A Semantic Differential scale rates the *attitude* of participants towards certain statements using antonyms (e.g. helpful-unhelpful). Where the Likert scale measures agreement from 1 - 7, the Semantic Differential scale measures a participant's attitude regarding the antonyms used towards a question on a scale of -3 - 3. Although both rating scales allow for degrees of opinion, there are subtle differences between them (Rosala, 2020). Answering a Semantic Differential requires more cognitive effort because the points on the scale are unlabeled. However, participants have more freedom in answering because this scale allows for more cognitive flexibility. The Likert scale type questions are generally easier to answer for participants but can be affected by acquiescence and social desirability biases (Friborg et al., 2006). When comparing results generated with both scales literature suggests that even though results generated with Likert and Semantic Differential scales are often very similar they cannot be used interchangeably because the sensitivity of the scales probably

differs (Schibeci, 1982).

Since these scales measure slightly different things, transforming the data the way that was done in the evaluation has maybe changed the results themselves. Furthermore, the transformation made the questions used in the questionnaire not validated, since they were only validated with a semantic differential scale. This does not necessarily mean the results of the evaluation are invalid because the Likert and semantic differential scale have some overlap. It does mean, however, that the claim that the statistical results of these metrics were generated with validated questions cannot be made, and that there might be a slight distortion in the data due to the transformation done; weakening the conclusions drawn based on the results generated in the evaluation. This is however not the case for the self-competence measure, because the validated questionnaire self-competence originated from already used the Likert scale. In hindsight, this difference between the Likert scale and Semantic Differentials should have been considered. In future work, these limitations should be taken into account.

Participant diversity

The participant group in the evaluation was quite homogeneous, so results of the evaluation cannot be extrapolated to master students or beginning researchers because of the differences in wants and needs between these groups, let alone to groups outside of the HMI discipline. However, the homogeneity was intentional to reduce variability between participants, so differences between groups would be clearer with such a mall sample size. This choice leads to several possibilities for future research and improvement of the advice. First, based on the insights into the wants and needs of other user groups such as master students and researchers, different versions of the advice could be developed and evaluated with these groups, so a different version of the advice becomes available for each group. However, since there are also large variations in skills and knowledge within these user groups, a second possibility is to make versions of the advice with different levels of helpfulness and give users the opportunity to choose which one they want to use based on their own needs. Then, the conclusions drawn about which advice works best for which group can be used to develop and evaluate advice for possible user groups in neighboring fields, or at different universities since it is unclear to what extent the insights gained about HMI at the University of Twente translate into other fields, universities, and countries. The research towards practices around ethics assessment in HMI already provided some implications that, even though the ethics assessment process in neighboring fields at the University of Twente, such as computer science, addresses different subjects, similar issues occur around it.

Tool questionnaires and advice

The current advice is not comprehensive and not checked for correctness. For practical purposes, advice within the tool does not cover the full range of topics present in the real ethics assessment procedure. For example, advice about doing research with participants who are low literate or speak little Dutch and English was not included in the advice, even though this is a type of typical research in HMI. Additionally, The correctness of the advice currently provided by the tool also leaves room for improvement since it was written by the author of this thesis who is quite aware of the struggles around the ethics assessment procedure but is not an ethics expert. However, if some promising results can already be achieved with this version of the advice, a more comprehensive version of advice may achieve very meaningful results. Therefore, Before the tool is implemented, the advice should be both checked by an ethics expert and the questions and extended to include all topics present in the ethics assessment procedure. The information provided about typical HMI research in Appendix F could be used as a starting point for this, but it may also be useful to look at the Oxfam flowchart from the previous section to have a different kind of example of how advice could be presented.

When the advice and questionnaires are improved, it is advised to also take a look at the advice questionnaire itself. Now, all questions are yes/no questions. These were fine for the evaluation

in which only a limited amount of topics were addressed and a case description that was made to fall within the questions asked in the advice questionnaires. However, if all topics within the ethics assessment procedure are included, the questionnaire might become too long, and yes/no questions are probably not fully applicable to actual research proposals. Therefore, it would be beneficial to look into what type of questions provide the right information from the user without them having to answer way too many specific yes/no questions. Maybe scenario-based questions describing typical research could already capture the core ethical issues of a significant part of the research proposals, while the yes/no questions can be used to determine ethical issues when these typical scenarios do not apply. The Eneri Decision Tree discussed in the previous paragraph provides an example of how questions could be asked differently, which could help inspire better ways to probe the user for information about their situation and what kind of information they might need.

Use of the tool by students and researchers outside HMI

During this thesis, a conscious decision was made during this thesis to limit research to within the University of Twente and focus on the HMI faculty. This left ethics assessment practices at other universities undiscussed and the advice in the tool unsuited for students and researchers outside HMI. In future research, it could be valuable to take a look at how other universities advise their students and researchers about the ethics assessment procedure. Maybe, there are certain practices already implemented there, which we could learn from. Also, they may encounter the same challenges, but try out different solutions. For example, the TU Delft provides a way to submit for minimal risk assessment, which suggests the challenges around research with minimal risks, as identified to be present at HMI during interviews, are also present there (Technische Universiteit Delft, 2022). However, there are probably also many differences between universities. Research into practices at other faculties and universities could help to get a better understanding of challenges around ethics assessment, and insights may eventually help extend the use of the tool beyond the HMI faculty.

Related work that was not considered

Apart from limitations and suggestions for future work regarding the evaluation, there are also some already existing projects to improve ethics assessment left unconsidered in this thesis. First of all, the background contains extensive information about the situation around research ethics and ethics assessment in HCI but does not contain any information about related work. This is because when this part of the report was written, the author was not aware of the existence of any assistance tools for preparing for the ethics assessment process, and was thus also not considered during this thesis. In the later stages of this thesis, however, some exciting tools popped up. These are shortly described here in relation to the tool used in this thesis, so they can be considered in future work if deemed necessary.

Eneri decision tree: The Eneri Decision Tree is an interactive PDF document that helps researchers to think through ethical questions and challenges they encounter (ENERI, 2020). It starts on the first page with very general subjects. Once the user selects what applies to them, the document skips to the page with the next question. Every question helps the user to further narrow down what kind of information they need. At the end of the decision tree, the document skips to the page with the information that applies to the question of the user and the points they probably need to consider for their research. The method of providing the right information and assistance through questions that narrow down the user challenges is very similar to the assistance tool used in this thesis. However, the target group, and therefore thy way information is presented, is different. The Eneri Decision Tree seems to focus more on members of ethics committees and researchers dealing with the research ethics around larger projects. The information provided is very complete and clear but is likely a bit overwhelming for a student trying to get information on how to answer questions from the self-assessment form of the ethics committee. The way they ask questions to narrow down

on a researcher's challenges, however, is very interesting. Questions are ordered as a list with subjects or examples from which the user can select what fits their situation best. Answering them feels more intuitive than the yes/no questions in the advice questionnaires in this thesis. Implementing these types of questions in the advice questionnaires in future iterations could make the questionnaires more intuitive.

Oxfam research ethics flowchart: Oxfam Great-Britain has published a research ethics flowchart in 2020 that provides a checklist of questions their researchers should consider in different stages of research and provides sources to official documents and guidelines (Mager & Galandini, 2020). Since this document is only focused on Oxfam UK the information it provides does not apply to the case considered in this thesis, but the way in which they present it is worth looking at. They have managed to present a summary of the points to consider in all stages of research and resources for further information in just two pages. Even the minimal advice designed in this thesis, which also mainly provided points to consider, was longer than that. They probably managed this by taking a lot of time to distill all key points their researchers should consider before, during, and after research and put all additional information in the documents the flowchart links to. This flowchart provides an interesting image of what the minimum advice could look like and might be worth considering for inspiration during the next iteration of the advice provided by advice questionnaires.

7.3 Conclusion

The goal of this thesis was to explore the problematic points of the ethics process in detail to be able to determine how to provide the right information at the right time through an existing web tool and determine what type of advice was most helpful, without the advice being so guiding that the sense of ownership over the prepared materials would deteriorate. From this objective, three research questions were formulated and answered during this thesis.

RQ1 - At which specific points in the process of gaining ethics approval do researchers struggle and why? From the research towards practices around ethics assessment in HMI, it can be concluded that (beginning) researchers mainly struggle with writing consent forms, information brochures, and other materials, identifying possible ethical issues, and applying feedback. This is not only caused by (some) disinterest in the process on their side but also by guidelines that are vague, confusing, too general, and too complicated to fully comprehend.

RQ2 – How can support on these specific points be implemented in the available tool while balancing helpfulness to the researcher, and the amount of learning and ownership of the researcher? Support on these specific points as proposed in this thesis consist of three different versions of advice questionnaires that provide advice applicable to the type of research the user wants to conduct. The answers given in the advice questionnaire determine which topics are included in the advice. All versions provide advice at different levels of helpfulness. The least helpful one provides only short points the user should not forget to include when preparing materials for the ethics assessment. The medium helpful one provides explanations of why certain topics need to be included and examples of how this could be done. the maximum advice version provides complete templates of needed forms and instructs users on how to prepare the necessary materials.

RQ3 – What is the effectiveness and impact of the resulting tool? The evaluation of the three versions showed that an advice questionnaire that provides tailored advice to the user can improve the self-competence of users in going through the ethics assessment procedure, and implied that it could improve the quality of the materials prepared for ethics assessment. For the user group that participated in the evaluation, inexperienced users, the medium advice performed best on all metrics, except for dependability. Also, the advice presented in the form of examples applicable to the user's planned research seemed to be most helpful. However, the interviews conducted to answer RQ1

already indicated that the type of advice that works best varies greatly between, and even within, user groups.

All in all, this thesis has shown a promising way of improving the preparation of materials for the ethics assessment procedure. However, before it can be implemented in HMI or other faculties, more research is needed towards what kind of advice is most useful for these other user groups, and the advice given needs to be extended to all subjects present in the self-assessment questionnaire.

References

Abascal, J., & Azevedo, L. (2007). Fundamentals of inclusive HCI design. In *International conference on universal access in human-computer interaction* (pp. 3–9).

A. de Lange, G. H. (2021). *Presentation ethics assessment tool.* (accessed on: 07/06/2021)

Branco, R. M., Quental, J., & Ribeiro, Ó. (2017). Personalised participation: an approach to involve people with dementia and their families in a participatory design project. *CoDesign*, *13*(2), 127–143.

Brown, B., Weilenmann, A., McMillan, D., & Lampinen, A. (2016). Five provocations for ethical HCI research. In *Proceedings of the 2016 chi conference on human factors in computing systems* (pp. 852–863). New York, NY, USA: ACM. doi: 10.1145/2858036.2858313

Bruckman, A. (2014). Research ethics and HCI. In *Ways of knowing in HCI* (pp. 449–468). Springer.

Calvo, R. A., Peters, D., Huppert, J., & Goggin, G. (2018). HCI as social policy: perspectives on digital rights in ethical design. In *Proceedings of the 30th australian conference on computer-human interaction* (pp. 74–82).

Cooper, A. K., & Coetzee, S. (2020). On the ethics of using publicly-available data. In M. Hattingh, M. Matthee, H. Smuts, I. Pappas, Y. K. Dwivedi, & M. Mäntymäki (Eds.), *Responsible design, implementation and use of information and communication technology* (pp. 159–171). Cham: Springer International Publishing.

CSDT. (2022a). *PCASS scales.* Retrieved from `https://selfdeterminationtheory.org/pcass/` (accessed on: 22/07/2021)

CSDT. (2022b). *Perceived Competence Scales.* Retrieved from `https://selfdeterminationtheory.org/perceived-competence-scales/` (accessed on: 21/07/2021)

Danish Board of Technology, T. (2015). *Satori factsheet.* Retrieved from `https://satoriproject.eu/the-project/` (accessed on: 14/02/2021)

Davis, B., Whitfield, C., & Anwar, M. (2018). Ethical and privacy considerations in cybersecurity. In *2018 16th annual conference on privacy, security and trust (pst)* (pp. 1–2).

ENERI. (2020). *ENERI decision Tree.* Retrieved from `https://eneri.eu/decision-tree/` (accessed on: 19/01/2022)

European Commision. (2018). *Data protection in the EU.* Retrieved from `https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en`

Frauenberger, C., Rauhala, M., & Fitzpatrick, G. (2016, 06). In-Action Ethics. *Interacting with Computers*, *29*(2), 220-236. Retrieved from `https://doi.org/10.1093/iwc/iww024` doi: 10.1093/iwc/iww024

Friborg, O., Martinussen, M., & Rosenvinge, J. H. (2006). Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Personality and Individual Differences*, *40*(5), 873–884.

Furby, L. (1978). Possession in humans: An exploratory study of its meaning and motivation. *Social Behavior and Personality: an international journal*, *6*(1), 49–65.

Gallagher, S. (2012). Multiple aspects in the sense of agency. *New ideas in psychology*, *30*(1), 15–31.

Goodyear, M. D. E., Krleza-Jeric, K., & Lemmens, T. (2007). The declaration of helsinki. *BMJ*, *335*(7621), 624–625. doi: 10.1136/bmj.39339.610000.BE

Guillemin, M., & Gillam, L. (2004). Ethics, reflexivity, and "ethically important moments" in research. *Qualitative inquiry*, *10*(2), 261–280.

Haggerty, K. (2004, December). Ethics creep: Governing social science research in the name of ethics. *Qualitative Sociology*, *27*, 391-414. doi: 10.1023/b:quas.0000049239.15922.a3

Jääskeläinen, R. (2010). Think-aloud protocol. *Handbook of translation studies*, *1*, 371–374.

Jamshed, S. (2014). Qualitative research method-interviewing and observation. *Journal of basic and clinical pharmacy*, *5*(4), 87.

Jansen, P., Reijers, W., Douglas, D., Sattarov, F., Gurzawska, A., Kapeller, A., ... Braun, R. (2016). A reasoned proposal for shared approaches to ethics assessment in the european context. *SATORI*.

KPMG. (2017). *Vernieuwde Declaration of Geneva goedgekeurd door WMA*. Retrieved from `https://www.knmg.nl/advies-richtlijnen/knmg-publicaties/https://www.knmg.nl/advies-richtlijnen/actualiteit-opinie/nieuws/nieuwsbericht/vernieuwde-declaration-of-geneva-goedgekeurd-door-wma.html` (accessed on: 22/01/2021)

Luger, E., & Rodden, T. (2013). An informed view on consent for ubicomp. In *Proceedings of the 2013 acm international joint conference on pervasive and ubiquitous computing* (pp. 529–538). New York, NY, USA: ACM. doi: 10.1145/2493432.2493446

Macnish, K. (2018). Government surveillance and why defining privacy matters in a post-snowden world. *Journal of Applied Philosophy*, *35*(2), 417–432.

Mager, F., & Galandini, S. (2020). Research ethics: A practical guide.

Molich, R., Laurel, B., Snyder, C., Quesenbery, W., & Wilson, C. E. (2001). Ethics in HCI. In *Chi '01 extended abstracts on human factors in computing systems* (pp. 217–218). New York, NY, USA: ACM. doi: 10.1145/634067.634197

Moore, A. D. (2008). Defining privacy. *Journal of Social Philosophy*, *39*(3), 411–428.

Moore, J. W. (2016). What is the sense of agency and why does it matter? *Frontiers in psychology*, *7*, 1272.

Moreno, M. A., Goniu, N., Moreno, P. S., & Diekema, D. (2013). Ethics of social media research: common concerns and practical considerations. *Cyberpsychology, behavior, and social networking*, *16*(9), 708–713.

Munteanu, C., Molyneaux, H., Moncur, W., Romero, M., O'Donnell, S., & Vines, J. (2015). Situational ethics: Re-thinking approaches to formal ethics requirements for human-computer interaction. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (pp. 105–114). New York, NY, USA: ACM. doi: 10.1145/2702123.2702481

Munteanu, C., Waycott, J., & McNaney, R. (2020). Dealing with ethical challenges in HCI fieldwork. In *Extended abstracts of the 2020 chi conference on human factors in computing systems* (pp. 1–4).

Nathan, L. P., Thieme, A., Tatar, D., & Branham, S. (2016). Disruptions, dilemmas and paradoxes: Ethical matter (s) in design research. *Interacting with Computers*, 1–9.

Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2015). Conducting semi-structured interviews. *Handbook of practical program evaluation*, *492*.

Pierce, J. L., Kostova, T., & Dirks, K. T. (2001). Toward a theory of psychological ownership in organizations. *The Academy of Management Review*, *26*(2), 298–310. Retrieved from `http://www.jstor.org/stable/259124`

Regulation, G. D. P. (2016). Regulation eu 2016/679 of the european parliament and of the council of 27 april 2016. *Official Journal of the European Union. Available at: http://ec. europa. eu/justice/data-protection/reform/files/regulation_oj_en. pdf (accessed 20 September 2017)*.

Resnik, D. (2021, 06). *Bioethics.* Retrieved from `https://www.niehs.nih.gov/research/resources/bioethics/` (accessed on: 19/01/2022)

Rosala, M. (2020). *Rating Scales in UX Research: Likert or Semantic Differential?* Retrieved from `https://www.nngroup.com/articles/rating-scales/` (accessed on: 19/01/2022)

Ruyter, K. (2019). The history of research ethics. *Forskningsetikk: Beskyttelse av enkeltpersoner og samfunn*, 17-25. Retrieved from `https://www.forskningsetikk.no/en/resources/the-research-ethics-library/systhematic-and-historical-perspectives/the-history-of-research-ethics/`

Sanderson, L. (2010). Informed consent in educational settings and the novice researcher. *Kairaranga*, *11*(1), 50–55.

Scharff, R. C., & Dusek, V. (2013). *Philosophy of technology: the technological condition: an anthology*. John Wiley & Sons.

Schibeci, R. (1982). Measuring student attitudes: Semantic differential or likert instruments?. *Science Education*, *66*(4), 565–70.

Schrepp, M. (2015, 09). *User experience questionnaire handbook.* doi: 10.13140/RG.2.1.2815.0245

Sims, J. M. (2010). A brief review of the belmont report. *Dimensions of critical care nursing*, *29*(4), 173–174.

Sudore, R. L., Landefeld, C. S., Williams, B. A., Barnes, D. E., Lindquist, K., & Schillinger, D. (2006). Use of a modified informed consent process among vulnerable patients: a descriptive study. *Journal of general internal medicine*, *21*(8), 867–873.

Technische Universiteit Delft. (2022). *Human Research Ethics.* Retrieved from `https://www.tudelft.nl/over-tu-delft/strategie/integriteitsbeleid/human-research-ethics/application` (accessed on: 19/01/2022)

Tribunals, I.-X. N. M. (1949). *Trials of war criminals before the nuremberg military tribunals under control council law no. 10.*

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and teacher education*, *17*(7), 783–805.

University of twente. (2021). *Bms ethical review.* Retrieved from `https://webapps.utwente.nl/ethicalreview/` (accessed on: 22/01/2021)

van der Burg, S., & van de Poel, I. R. (2005). Teaching ethics and technology with Agora, an electronic tool. *Science and Engineering Ethics*, *11*(2), 277–298. doi: 10.1007/s11948-005-0046-8

Vitak, J., Proferes, N., Shilton, K., & Ashktorab, Z. (2017). Ethics regulation in social computing research: Examining the role of institutional review boards. *Journal of Empirical Research on Human Research Ethics*, *12*(5), 372–382.

Waern, A. (2016). The ethics of unaware participation in public interventions. In *proceedings of the 2016 chi conference on human factors in computing systems* (pp. 803–814).

Waycott, J., Munteanu, C., Davis, H., Thieme, A., Branham, S., Moncur, W., ... Vines, J. (2017). Ethical encounters in HCI: Implications for research in sensitive settings. In *Proceedings of the 2017 chi conference extended abstracts on human factors in computing systems* (pp. 518–525). New York, NY, USA: ACM. doi: 10.1145/3027063.3027089

Wendler, D. (1996). Deception in medical and behavioral research: is it ever acceptable? *The Milbank Quarterly*, 87–114.

Wet op het bevolkingsonderzoek - bwbr0005699. (1998, 01). Retrieved from `https://wetten.overheid.nl/BWBR0005699/1998-01-01`

A Matrix of ethics assessment process

This matrix contains the steps of the ethics assessment procedure combined with the main question relevant to the goal of the research. These are displayed on the y-axis of the matrix. The other axis contains the different perspectives that will be addressed. The selection of relevant steps for each group is indicated in the first few columns of the matrix.

process step						construct	Goal/output	Obstruction
1. Recognition of need for ethics assessment						Do researchers recognize the need for ethics assessment	awareness of the need for ethical assesment	Forgetting or not knowing about it
Identification of necessary actions						Are researchers able to find out which steps they need to take for getting ethical approval	List of actions	Not knowing where to find information
Preparation of materials						Are researchers able to prepare the necessary materials identified in the previous step	Consent form, information brochure, completed checklist, GDPR consent	
- consent form							Consent form	Does not know how to draft this
- information brochure							information brochure	Does not know how to draft this
- Data recording and analysis consent (GDPR)							GDPR form	Does not know how to draft this
Apply for approval						Do researchers know who can grant approval and how to reach them	Mail containing all materials to ethics committee	Unawareness of how to reach ethics committee
Review proposal						Are the criteria of review unambiguous and is review objective	Criteria on which proposal is assessed	Unclear guidelines, differences in interpetation
Review of proposal								
Getting feedback						Is the feedback recieved on the proposal clearly communicated	Feedback to researcher	Late replies, unclear phrasing
Applying feedback						Are researchers able to apply the recieved feedback properly	Improved proposal	Lack of abilities, quality of feedback
Informing participant/get consent						Is the goal of informing the participant to allow them to make a considered choice reached	Informed consent participant	Uninterested participant, clarity of information brochure, unhelpful researcher
Do research						Is the researcher able to deal with unexpected ethical decision making during research	Ethical decisions during research	Lack of training, lack of percieved importance, time pressure
Data analysis						Does the data analysis still fit the plan made beforehand in the ethical assessment	Analyzed data	Unexpected findings, change in research direction
Data storage						Does the researcher comply to the way of data storage described in the ethical assessment beforehand	Secure data storage	Carelesness, available storage space, forgetfulness

Figure 12: First part of matrix. Colours identify which participant group is interviewed on a part of the process. LTR: First and second year students, end of Bachelor/ start of Master, experienced Master students, PhD and researchers, ethics committee members.

process step	possible sourses of help	Knowledge/abilities	Efficacy	Percieved importance
1. Recognition of need for ethics assessment	Peers, supervisors and teachers	Awareness of the importance of ethics assessment	NA	NA
Identification of necessary actions	Peers, guidelines, teachers, ethics committee	Knowing which steps you need to take to get ethical approval	Feeling like you know all necessary actions	NA
Preparation of materials	Peers supervisors, teachers, guidelines, ethics committee	Knowing how to draft these texts	Feeling capable of succeeding in this	
- consent form	Peers, guidelines, teachers, ethics committee	Ability to formulate the things a participant has to consent to as clearly as possible given the participants needs	Feeling capable of doing what was discribed in the previous section	
- information brochure	Peers, guidelines, teachers, ethics committee	Ability to determine all key features of the research and communicate this in a way that fits the participants needs	Feeling capable of doing what was discribed in the previous section	
- Data recording and analysis consent (GDPR)	Peers, guidelines, teachers, ethics committee	Ability to communicate what happens with the participants data after research and what their rights are in a way that fits the participants needs	Feeling capable of doing what was discribed in the previous section	
Apply for approval	Peers supervisors, teachers, guidelines, ethics committee	Knowing about the ethics committee and what they do	NA	NA
Review proposal	Guidelines, other members ethics committee	Agreement about application of regulations during review	Feeling capable of applying regulations and assessing proposals objectively	
Review of proposal				
Getting feedback	Ethics committee, teachers, supervisors	Ability to give clear feedback of the ethics committee	percievedc ability to give clear feedback	NA
Applying feedback	Peers, supervisors, teachers, guidelines, ethics committee	Knowledge abuot guidelines to application when necessary	Feeling able to deal with feedback and make changes	
Informing participant/get consent	Supervisors, teachers	Knowing the level of understanding of the participant and adapt to that. Providing a short and simple explanation of research containing all key features.	feeling comfortable and able to communicate all key parts of the research to the participant	
Do research	Peers, supervisors, teachers, guidelines, ethics committee	Knowledge about ethical behaviour and theory to support decisions	Feeling able to make an ethical decision when the situation calls for it	NA
Data analysis	supervisors, teachers, guidelines, ethics committee	Being able to assess when data analysis changes fall within or outside the consent given by the participant	feeling able to assess this and make a good choice if the sort of analysis is allowed or not	
Data storage	supervisors	Being compliant with data storage as discribed in the GDPR form	NA	

Figure 13: Second part of matrix

B Examples form interview guide

Perspectives are color coded. green = knowledge/abilities, purple = sources of help, blue = self assessed ability, red = obstructions, yellow = perceived importance

Statements

Percieved importance of ethics assessment

- Ethics assessment in HCI is just as important as in other fields like medicine and sociology.
- Ethics assessment is not that important for the type of research I conduct.
- I would not want to conduct research if the ethical implications of my research were not reviewed by an objective committee.

Self assessed ability to go through process in general

- I have a good understanding of what is ethical research and what is not
- I am able to complete an ethical assessment request without any assistance
- I always ask someone to help me with going through the ethical assessment process
- I often feel lost or frustrated when going through the ethical assessment process

Willingness to go through the process

- If there was a way to avoid the ethical assessment process of my research experiments I would always avoid it.
- I don't mind going through the ethical process and filling in the ethics checklist
- Going through the ethics process is something I delay doing

Interest in ethics

- Going through the ethics procedure feels like a chore
- Ethics of research is an interesting topic to have a discussion about
- I like thinking about the ethical implications of decisions, research or new technologies
- I never think about ethics in any way outside of the ethics assessment procedure

Ethics assessment (this is ethical - this is ethically questionable - this is unethical)

- Dilemma 1: You want to test an application for illiterate adults, so you let them sign a consent form for this that you first read aloud. Some participant says you should not bother explaining. They really want to participate because your tool can make it easier for them to learn basic reading skills. They say they trust you and sign the form without knowing its contents. You still let them participate.
- Dilemma 2: A participant of the study in the previous dilemma has tested the tool over the weekend and tells you that his wife, who is not part of the study but is also low literate, has played around with the tool and has told him about it. Her experiences yield some very valuable insights into improving the tool. You decide to include her feedback in the next iteration.
- Dilemma 3: You are doing a study with electronic toys in a primary school. Some children did not hand in the consent form their parents had to sign because they forgot it at home. They are really looking forward to playing with the toys, So you still let them take part in the experiment.
- Dilemma 4: The situation is the same as above, but in this case you let them play with the toys in a separate group, but do not record any data of these children.
- Dilemma 5: You are developing a special controller for handicapped people, so they can play the same games as their friends. During testing of some prototypes however, it is likely they will experience temporary joint pain because your ergonomics aren’t great.

You mention this to your participants, but this does not deter them from wanting to try out the new controller.

Interview questions per topic

Do researchers recognize the need for ethics assessment

- Can you explain when ethics assessment is necessary in your own words?
- If you are not sure if ethics assessment is necessary, who would you ask/ what would you check?

Are researchers able to find out which steps they need to take for getting ethical approval

- Do you think you would be able to find out what steps you need to take to get ethics approval?
- Suppose you found out that it is necessary to get approval from the ethics committee, What are the steps? (If you do not know them), how would you go about finding out what steps you need to take to get this?

Are researchers able to prepare the necessary materials identified in the previous step

- Do you know what forms or materials are necessary for ethics approval? If not, who would you ask for help or advice?
- Do you use examples of the necessary forms? Where do you find them?
- Would you feel confident that you would manage to deliver all these materials?
- How important are consent forms, data usage agreement and information brochures given the research you engage in? Why do you think that?

Do researchers know who can grant approval and how to reach them

- Who is responsible for granting approval of the ethics assessment of your research? How would you contact them?
- If not, do you think you can find this information yourself ? Who would you ask to get this information?

Is the feedback received on the proposal clearly communicated

- How quickly do you receive feedback from the ethics committee? And in what form?
- Do you understand the feedback given and why it is given?
- Do you feel like you can approach the committee if anything is unclear? Why(not)?

Are researchers able to apply the received feedback properly

- Is it clear from the feedback what improvements you need to do on your proposal?
- Are you able to do these improvements yourself? If not, what is not clear, who do you ask or where do you look up information?
- Do you feel confident that your proposal will be accepted after you have made the requested changes?

- How crucial do you think the changes you make are for the quality of ethical assessment?

Is the goal of informing the participant to allow them to make a considered choice reached

- Can you describe how you go about informing a participant before they give consent?
- Do you adjust the way you inform a participant to the person? In what way?
- Do participants ask questions? What kind of questions?
- What do you think participants base their decision to give consent on?
- What is the value of the current process of informed consent in your opinion, given the type of research you generally do?

Is the researcher able to deal with unexpected ethical decision making during research

- Have you ever been in the situation where something unexpected happened during research? In what way did this affect you and your research? Other questions only useful if this is the case:
- Looking back at this situation, would you have liked to handle things differently?
- Do you feel like you are always prepared for unexpected difficulties that might occur during research? Why (not)?

Does the data analysis still fit the plan made beforehand in the ethical assessment

- Are findings from the data collected generally as expected? Has the research goals ever changed based on preliminary findings?
- If yes, did the process of data analysis chance as a result of that? Was it necessary/important to inform participants about these changes, and why?
- If no, what would you do if the data collected is different than expected?
- Do you feel able to assess if a slight change in data analysis falls within or outside the consent given by participants?

Does the researcher comply to the way of data storage described in the ethical assessment beforehand

- Where do you store data collected during research?
- What do you do with consent forms and data once research is finished?
- Has any participant ever requested the deletion of their data?
- Are there cases in which you could not/did not adhere to the data storage requirements as described in the Ethical assessment? Why?

Final question

- Are there any particular things in the ethics assessment process you think could be improved?

C Processed and combined answers ethics committee members

Questions and answers are denoted with the (Q) and (A) respectively. **Boldface** indicates a certain step in the ethics assessment process. *Italics* indicates an observation or interpretation of the researcher. A number with () indicates how many of the 6 people have mentioned a certain statement.

(Q) Current activities at the university of twente (A) One with training in ethics in the university wide committee, most people who are part of the committee have experience in their field but no official training.

(Q)For how long have you been in an ethics committee? (A) All members seem to be quite new to it which is understandable since in many fields where there is not a lot of research done with human participants, ethics assessment itself is quite new. Max is 2 years, the newest members have joined a couple of months ago.

(Q) Why did you decide to join the committee? (A) All were asked by others already part of a committee. One was asked because of his formal training. Others had shown interest in or experience with ethics within their field and were asked because of this. One member was asked because he had experience with setting up an ethics committee for computer science.

(Q) Why do you think ethics assessment is important? (A) According to the philosopher: important because: 1.establishing social trust, social legitimacy, strengthened trust if scientists care about ethics 2.justified consequentially;engaging in ethics gives better outcomes 3. intrinsic grounds; feels important to do. These reasons were also mentioned by others in other words. Point 1. Was mentioned by 2 others, point 2 and 3 were mentioned by all others in some form or another.

Elaboration on point 2: Engaging in ethics gives better outcomes because; it makes you think about your protocol, It makes you think about how to handle vulnerabilities, makes you think about possible implications of the outcomes of your research for society

(Q) Do you enjoy your work in the committee? Why or why not? (A) All like the work for generally the same reasons; interest, the feeling that they add value/do something important, nice to see what others work on in different research groups. But it also has its downsides. Downsides mentioned are; the time it takes up (4), other researchers do not seem to recognize the importance of the ethics committee/lack of recognition(2), researchers have other interests and little ethics training which makes discussing issues difficult (1), large bias in the direction of letting people go through with their research(1)

Do researchers recognize the need for ethics assessment (Q) What is the attitude of researchers towards ethics assessment and the ethics committee? What is the cause of this in your opinion? (A) Generally positive, people seem willing to improve their research(3). However, people do not think about ethics regarding their research in research groups where there is little participant research(3) some more details on this: Struggles are in every field but stronger in engineering etc. because people are less trained in ethics and less exposed to this way of thinking at a technical university. This combined with the fact that possible societal implications of technology can be quite large and influential makes it difficult for engineers to be motivated and or engage seriously in ethics. Also, incentive of research money and promotions motivates to do research even if the ethics might be complicated(1). If you point out ethical issues people do take it seriously and engage with it (3). Very little appetite for telling people that the cannot do things. Attitudes seem to be more towards "help people do their research ethically" instead of "can we do this research given our ethical standpoints" People do not like hard restrictions(2) other perspective on this: Disposition of not wanting to say no is stronger in engineering etc. Is aware that he himself does not have to pay the costs if he says no we cannot do this research. At committee only 2 people with formal ethics training(1). People do not like spending a lot of time on ethical assessment because they are already busy, or in a hurry to start

their research.(3)

From the conversations in this question I gathered that people are willing to engage in ethics but have little experience in fields where there is little participant research, are afraid ethics assessment will prohibit them from doing the research they want to do and think it is a lot of work. However, if ethical issues are pointed out to them they do engage with them to try to solve them and seem very willing to improve their research. The biggest problem here seems to be a lack of experience. People have a lot of preconceptions about the ethics assessment process, but if they engage in discussion with a committee member they do engage in ethical issues properly and sometimes even with interest. With one or two exceptions.

Are researchers able to find out which steps they need to take for getting ethical approval (Q) Do you feel like researchers are generally aware of the actions they need to take to get ethics approval? (A) They are generally aware of who to talk to, but the questions they have to answer are seen as unclear and supervisors have to point out that they still have to do the ethics assessment. Students are not pro-active. Also in addressing ethical issues this is missing. People are reactive, also PHDers and researchers. People seem to miss an overview of what needs to be done.

(Q) What do you think about current process steps researchers have to follow? Any suggestions on possible improvements in this regard? (A) Improvements are already being made with better checklists etc. so the process is changing. Things that may still be missing is more holistic ethics training to properly engage with the process, give people completer examples, (for computer science) a protocol for what to do when you discover a vulnerability in a system.

Are researchers able to prepare the necessary materials identified in the previous step (Q) What do you think of the quality of the material that is submitted to the committee? What could be improved in your opinion? (A) This is very mixed. For social sciences quality is okay especially after one review. The quality is improving since supervisors are pointing out problems before the request is sent to the committee. In computer science/HMI/Creative Technology the quality is very mixed and is very dependent on the individual. People that are experienced and/or interested in ethical assessment do quite good. Others do terrible and miss entire ethical issues. Quality in data science specifically is terrible because people do not recognise ethical issues. If they are pointed out to them they do relatively okay.

(Q) Can you think of any specific mistakes that are made very often? (A) Most mistakes are do to missing (indirect) ethical issues or forgetting to explain thing. Most mistakes are do to missing (indirect) ethical issues or forgetting to explain things properly. One specific mistake mentioned for computer science: People focus on all the risks and stating risks instead of focusing on weighing risks/rewards. Risks are fine as long as the benefits in terms of research are much higher. Weighing these happens mostly after issues are pointed out Do researchers know who can grant approval and how to reach them (Q) What kind of assistance do you provide to researchers? (A) Most assistance consists of pointing out dilemmas they did not address, explaining dilemmas to them and pointing out where they missed some explanation or mistake. Occasionally discuss about the research protocol and how to still get the desired data in a way that is ethical. Are the criteria of review unambiguous and is review objective (Q) Do you have the feeling ethics boards are able to apply the current regulations consistently? (A) Yes, regulations are clear to everyone and there are never discussion about the regulations themselves

(Q) Do you ever have discussions within the committee about accepting or rejecting a proposal? What are the topics (A) of these discussions generally? How do you resolve this? This almost never happens, the only discussions that happen are about interesting cases that are ethically questionable or complicated. But almost all requests are straightforward. Also, pre checkers sometimes point out ethical issues to the committee so they know what to focus on.

Is the feedback received on the proposal clearly communicated (Q) How long does it take to get back to someone with feedback? (A) Generally between 1-2 weeks,also depends on how quickly the student/researcher responds. but in some extreme cases which take a lot of mailing back and forth it takes a month.

(Q) Do you get a lot of questions about the feedback you give? (A) Questions happen mostly before the proposal is sent to the committee at the pre checkers. The feedback almost never leads to questions. It seems clear what people need to change. Are researchers able to apply the received feedback properly (Q) If a proposal gets rejected, how well are improvements generally done when you receive it a second time? (A) For social sciences: 2/5 no comments 2/5 one round of improvement 1/5 multiple rounds of improvement In data science: they will do the work and make relatively good adjustments when ethical problems are pointed out In Creative Technology/hmi/computer science: Most people fix issues in one round of feedback. generally multiple rounds are needed only when feedback concerns an ethical issue. People do seem very capable in applying changes once they understand the ethical issue but this can take a bit of discussion. most problems are already taken care of before it goes to the committee by the pre-checkers.

(Q) Final question: any ideas for improvements on the ethics process?

- University wide; clarifying what the mandate of the ethics committee is. More communication and backup from the executive board. Unclear if they will support a decision.
- BETTER TRAINING (this was stressed multiple times) on different topics. 1) can you fill in a good ethics part for a grant 2) research integrity 3) knowing the ethical consequences of your research. All important and different. The amount of training students get at different subjects seems to be disproportionate. 30 seems to be neglected. It would be great to have a coherent system for ethics training of all students to make sure they are properly trained in all of them.
- Give students good examples of ethical implications in all the courses about AI. students that do engage in ethics had to do all the work themselves. Giving them awareness should be embedded in existing AI courses. If students are given good examples on every step of the way they will start doing this themselves
- first, training ethics assessors needs to be done.
- Procedure needs to be streamlined so everything can go quicker. The new electronic tool will help with that.
- Also there should be more awareness for ethics among researchers. They can then pass it on in their teaching
- Improve communication. Give everybody a case study in the bachelor so they have some experience.
- Furthermore, supervisors should also get more practice in this so they can better help to improve their students work. This could also help relieve some of the pressure on the pre-check committee
- create an Ethics advisor.
- Better protocol for unexpected discoveries. Cooperation between universities the way they do it in medicine.
- Give student concrete examples so they get used to the way of thinking about ethical issues of research

D Matrix with processed results interviews

D.1 Students

process step	construct	Goal/output	Obstruction
1. Recognition of need for ethics assessment blue = said by supervisors	Do researchers recognize the need for ethics assessment	awareness of the need for ethical assesment	unsure when personal data is gathered, however this was only mentioned once, students forget about it.
Identification of necessary actions	Are researchers able to find out which steps they need to take for getting ethical approval	List of actions	none, cannot find (the correct) guidelines, does not know what info is deeed for the process
Preparation of materials	Are researchers able to prepare the necessary materials identified in the previous step	Consent form, information brochure, completed checklist, GDPR consent	unclear questions in checklist (1), questios are vague or do not apply to the research
- consent form		Consent form	Unsure if they included everything that needs to be there (4), does not know where to start(1), unsure what it should contain
- information brochure		information brochure	Unsure if they included everything that needs to be there (4), does not know where to start(1), unsure what it should contain
- Data recording and analysis consent (GDPR)		GDPR form	Does not know it exists(2), filled in the wrong form (1)
Apply for approval	Do researchers know who can grant approval and how to reach them	Mail containing all materials to ethics committee	Does not know pre-checkers exists (2), never dealt with the commitee directly so no ecperience(1),
Review proposal	Are the criteria of review unambiguous and is review objective	Criteria on which proposal is assessed	Unclear guidelines, differences in interpetation
Review of proposal			
Getting feedback	Is the feedback recieved on the proposal clearly communicated	Feedback to researcher	It takes too long (3), feedback unclear(1)
Applying feedback	Are researchers able to apply the recieved feedback properly	Improved proposal	feedback unclear(1), feedback difficult to implement because of unknowns in research(1), stupid mistakes
Informing participant/get consent	Is the goal of informing the participant to allow them to make a considered choice reached	Informed consent participant	Participants who do not want to read a lot of text but sometimes a lot of info needs to be given (2)
Do research	Is the researcher able to deal with unexpected ethical decision making during research	Ethical decisions during research	none were mentioned
Data analysis	Does the data analysis still fit the plan made beforehand in the ethical assessment	Analyzed data	cange in order of tests (1), does not really help you any closer to the goal (2)
Data storage	Does the researcher comply to the way of data storage described in the ethical assessment beforehand	Secure data storage	carelesness(2), not being able to edid data in the way they want(1), too much work (1), not worthe the hassle for small research projects.

Figure 14: First part of matrix

process step	possible sourses of help	Knowledge/abilities
1. Recognition of need for ethics assessment blue = said by supervisors	Guidelines, collegues, Supervisors	Every participant was able to name one valid reason why ethics assessment is important. it is very dependent on the sudent. most do okay in recognizing the need
Identification of necessary actions	none, everybody can locate guidelines, supervisors	Everybody knows where to find guidelines and can name all main steps in the process. only the GDPR registration was forgotten twice and one person did nit name drafting the forms as a step. Again very dependent on student; some know morw than supersisor some really struggle
Preparation of materials	supervisor, peers	it is clear to participants what forms are needed but some questions in the checklist anre unclear to some (the once with the least experience) Experience makes all the difference. if they already have done it is is much simpler because tey already kind of know what to fill in
- consent form	dennis(6), supervisor(4), pre-checker(1), examples (7)	everybody knows what the consent form is, why it is needed but the exact contens of it is a bit unclear. see row 4
- information brochure	dennis(6), supervisor(4), pre-checker(1), examples (7)	Everybody knows what an information brochure should do, but only half of the participants could name the things it should contain in more detain than; "explanation of the research you are doing" see row 4
- Data recording and analysis consent (GDPR)	Ethics committee (1)	the people who know it exists were clear about how they handle this. just fill it in, it speaks for itself.
Apply for approval	Peers supervisors, teachers, guidelines, ethics committee	All participants know how to reach the ethic committee, but only six knew about the pre-checkers. They say this is because this is not specified on the website. (are there multiple versions of the site?)
Review proposal	Guidelines, other members ethics committee	Agreement about application of regulations during review
Review of proposal		
Getting feedback	Ethics committee member, dennis	Except for one case, feedback was clear but took somtimes a bit too long
Applying feedback	ethics committee, dennis,advice from supervisors	Most are able to apply feedbach well. the ones that were not say they were able after talking to an ethics committee member. genarally good but very person dependent, sometimes you have to make them understand what the problem is but if they get it they gelerally deal with it well
Informing participant/get consent	supervisor	Participants seem aware of the kind of participant they are dealing with and how to provide fitting information. almost all check if their forms have been read in some way before research starts. they do not get many questions from participants.
Do research	none were mentioned	Participant seem very capable in behaving ethically and dealing with unexpected situations if they were encoutered(1). This is not because of good training (according tho them) but becaus of generally good social skills of students
Data analysis	supervisors	Participants say that even if data is unhelpful or there is a slight change to the research methodology this is not problematic because the method does not change and the consent form is so general that a gange in analysis of aquired data does not require extra permission. If it would they would ask the participant.
Data storage	none were mentioned, providing lockers on university for forms	All participants are aware of the regulations. if they are not followed it is because of the reasons named in obstructions, they know what should be done but they do not always do it according to some supervisors. Students which do more complicated/sensitive research do better

Figure 15: Second part of matrix

process step	Efficacy	Percieved importance
1. Recognition of need for ethics assessment blue = said by supervisors	NA	NA
Identification of necessary actions	of the 8 participants, 2 were unsure about their ability to identify the steps. even though they were able to name the majority of steps. If it is one of there first times doing this it is low. higher if they have some experience	NA
Preparation of materials	generally good, see below for more info	preparation in general relatively mportant, because if you do it well it saves you time. Very dependent on student. some take it too seriously, some dont give a shit. this is reflected in quality of material
- consent form	all of the participants feels capable of drafting this. although most need an example for it and half lets their supervisor check it Generally this is true: good if they find the process important and found a good example, bad if they do not care	the combination of the conset for and infromation brochure were seen as important, but various reasons were named. because you want to have participants counsiously choose to participate, because it protects the researcher, because it informs the participant about their rights etc. see row 4
- information brochure	all of the participants feels capable of drafting this. although most need an example for it and half lets their supervisor check it Generally this is true: good if they find the process important and found a good example, bad if they do not care	see row 5. People see these as one package see row 4
- Data recording and analysis consent (GDPR)	The once who know about the form find it easy to fill in. The ones who do not know about it were suprized when i asked about it. the person who filled in the wrong form (one for a grant) said he will be able to find the right one next time.	Not important bacause it was only mentioned by one participant as something important. I think people see it as a burocratic nessesity.
Apply for approval	NA	NA
Review proposal	Feeling capable of applying regulations and assessing proposals objectively	relatively important. participants like that there is somebody checking these things. even though they might not feel like a check is that important for their own standard research.
Review of proposal		
Getting feedback	NA	NA
Applying feedback	Generally participants feel able to apply feedback or ask for help if they are unable. they are also confident that the proposal will be accepted after they have made changes. generally okay, but it is quite mixed. experience and interest play a role here	Not that important. changes are mostly minor and doe not influence the research procedure. depends on what the feedback is. unimportant with minor thing but if an ethical issue is pointed out the students that are interested in the process find this quite important
Informing participant/get consent	They give the impression of being confident in informing participants.	This step is seen as very important. reasons given are, protection of participant, providing cosnistent information, provide enough information so people can make their own choice
Do research	the participants all feel able to deal with unexpected situations if they might present themselves. only one of them has encoundered an unexpected situation with ethical implications so i cannot say if this is true	NA
Data analysis	All say they are able to assess if changes fall within the onsent given or not	They dont see it as important to notify participants about slight changes in analysis because it does not go beyond the consent given in their experience
Data storage	NA	If the data is kept in a place were participant sthink it is safe, some do not bother with the requirements for data storage. I get the feeling that this is also partly caused by the fact that data they gather is not that sensitive. THere just not that conserned about privacy. the facht that they are mostly doing standard research with little risk doesnt help. Students which do more complicated/sensitive research care more

Figure 16: Third part of matrix

D.2 Researchers

process step	construct	Goal/output	Obstruction
1. Recognition of need for ethics assessment	Do researchers recognize the need for ethics assessment	recognition of the need for ethics assessment	1. Researchers from fields related to HMI who do little/no reseach with participants (n=3) mentioned that it can be unclear for them if ethics assessment is needed when they analyze large data sets or do interviews with experts. 2. Researchers and PHD within HMI (n=2) mentioned not seeing the nessesity of such an elaborate procedure for standard research such as surveys and interviews.
Identification of necessary actions	Are researchers able to find out which steps they need to take for getting ethical approval	List of steps of the ethics assessment process	1. one researcher finds the rules and regulations unclear which gets in the way of identifying nessesary actions
Preparation of materials	Are researchers able to prepare the necessary materials identified in the previous step	Consent form, information brochure, completed checklist, GDPR consent	1. One researcher mentioned that preparing these materials according to the guidelines of the university feels like too elaborate for very low risk standard research such as interviews with experts. 2. Others did not metions any obstructions altough they did list sources of help
- consent form		Consent form	some are unsure if it contains the nessesary information
- information brochure		information brochure	some are unsure if it contains the nessesary information
- Data recording and analysis consent (GDPR)		GDPR form	none
Apply for approval	Do researchers know who can grant approval and how to reach them	Mail containing all materials to ethics committee	none
Review proposal	Are the criteria of review unambiguous and is review objective	Criteria on which proposal is assessed	Time constraints at the side of the committee (n=1)
Review of proposal			
Getting feedback	Is the feedback recieved on the proposal clearly communicated	Feedback to researcher	It takes too long to recieve approval from the committee itself (n=6)
Applying feedback	Are researchers able to apply the recieved feedback properly	Improved proposal	none
Informing participant/get consent	Is the goal of informing the participant to allow them to make a considered choice reached	Informed consent participant	1. some participants do not read the information brochure 2. expert participants feel like its unnessesary to give consent because they are engaged in the subject professionally. its their job! 3. dilemma between how to provide enough information in a format participants with difficulty reading/understanding still understant
Do research	Is the researcher able to deal with unexpected ethical decision making during research	Ethical decisions during research	lack of experience is a possible obstruction for beginning researchers there were different stories about "when i first started out..... but later on this problem never occured again because i antocipated it and took measures"
Data analysis	Does the data analysis still fit the plan made beforehand in the ethical assessment	Analyzed data	none
Data storage	Does the researcher comply to the way of data storage described in the ethical assessment beforehand	Secure data storage	encrypting everything can be complicated, you cannot work on analysis at the place/time you might want

Figure 17: First part of matrix

process step	possible sourses of help	Knowledge/abilities	Efficacy
1. Recognition of need for ethics assessment	Collegues (n=3)	All researchers apart from 2 on the fringes of HMI mentioned in obstructions gave the impression that they were very capable of recognizing the need for assessment.	NA
Identification of necessary actions	supervisors (n=2), presentation from old course (n=1)	all researchers except one were able to list all general steps of the process	all researchers except two feel confident about this. One of them has the problem mentioned in obstructions, the other does a lot of standard research and is not sure he would be
Preparation of materials	Examples from others (n=2), dennis (n=3), own examples (n=4)	all Participants -1 from this group are able to list all general steps in drafting these and name point of intnerest.	all feel capable of succeeding in this
- consent form	see row 4	all Participants -1 from this group are able to list all information that neest to be in there	all -1 feel capable of succeeding in this
- information brochure	see row 4	all Participants -1 from this group are able to list all information that neest to be in there	all feel capable of succeeding in this
- Data recording and analysis consent (GDPR)	the documentation provided(n=1)	only one participant was not aware of this (researcher) but that is probably because PHD ers fix everything. others seem able to do it	all feel capable of succeeding in this
Apply for approval	no needed	people know how to reach the committee	NA
Review proposal	none	everyone seemed to generally agree with the verdicts of the committee, although one mentioned they thought one of their proposals was not reviewed thouroghly enough once due to time constrains at the side of the committee.	NA
Review of proposal			
Getting feedback	dennis(n=2)	almost never feedback once the application is sent to the committee itself. feedback from pre-checkers/supervisors is seen as clear and useful	NA
Applying feedback	none	all participants seem to know how to deal with feedback	all participants feel like they are able to deal with feedback
Informing participant/get consent	include short explanation before consent is signed, involve caretakers etc. some organisations even have guidelines in place	In general the researchers who do a lot of research themselves still seem to have a very clear idea about how to properly inform different types of participants. experience is given as an iportant factor for this. Researchers who only do standard research have more difficulties with discribing how they would inform participants with diffeent skills.	Researchers are quite aware of their own skills. most have a feeling that they are able to inform people well and get consent in an ethical way. The researchers that do only standard research are aware of their lack of skills in working with vulnerable groups
Do research	supervisor	3 experienced unexpected situations. all of them were not very critical and they said they dealt with them in a good way.	researchers are confident that they can deal with unexpected situations and have the feeling that they can always find support with their collegues if not.
Data analysis	none	about half of the researchers has dealt with unexpected data. this has never lead to a change in research direction. it is more the case that can inspire a new research direction. the phrasing about which data is alalysed is always sufficiently broad that unexpected data is not a problem. One researcher went back to the participant to ask for extra permission, but this was for a quote.	everyone interviewed feels able to assess if a different way of analysing the data falls within the consent form
Data storage	none	even though some shared a mild frustration about the extreme caution with data, all researchers told be they adhere to the requirements (almost) all the time. They save it on an encrypted drive, university drive, in a locked storage space and distroy/delete at the required time	NA

Figure 18: Second part of matrix

process step	Percieved importance
1. Recognition of need for ethics assessment	NA
Identification of necessary actions	NA
Preparation of materials	The consensus is that these forms are always somewhat important, but become more important if research is risky or with specific (vulnerable) groups of participants.
- consent form	see row 4
- information brochure	see row 4
- Data recording and analysis consent (GDPR)	generally people see this as important but I have heard different resasons for this. 1. because its the law 2. to keep track of which data is still out there 3. you just have to. I do not get the feeling that people are very sure about why its needed but they still say they view it as important.
Apply for approval	NA
Review proposal	
Review of proposal	
Getting feedback	NA
Applying feedback	
Informing participant/get consent	in general it is percieved as important for the research they do by all participants - 2. the reasons 1. give people the chance to make their own (informed) decision to participate. give them agency (n=6) 2. protect participand and/or researcher (n=3) 3. give participants an idea of what they can expect (n=2) 4. give consistent information to improve the quality of your research (n=1). 5. it is not that important for the research i do (n=2) for this one was not working with personal data, the other only interviewd experts
Do research	NA
Data analysis	Everyone finds it important that the analysis is kept within the permissions of the consent form but since nobody ever had a problem with this, there is not much to say about this
Data storage	The importance people give to this seems very dependent on the nature of the data. Researchers who do standard research withconsenting adults do not stress the importance of being careful with data and were you save it. while researchers who do research with vulnerable participants/ sensitive data find correct handling of data very important.

Figure 19: Third part of matrix

E Full overview of editor environment

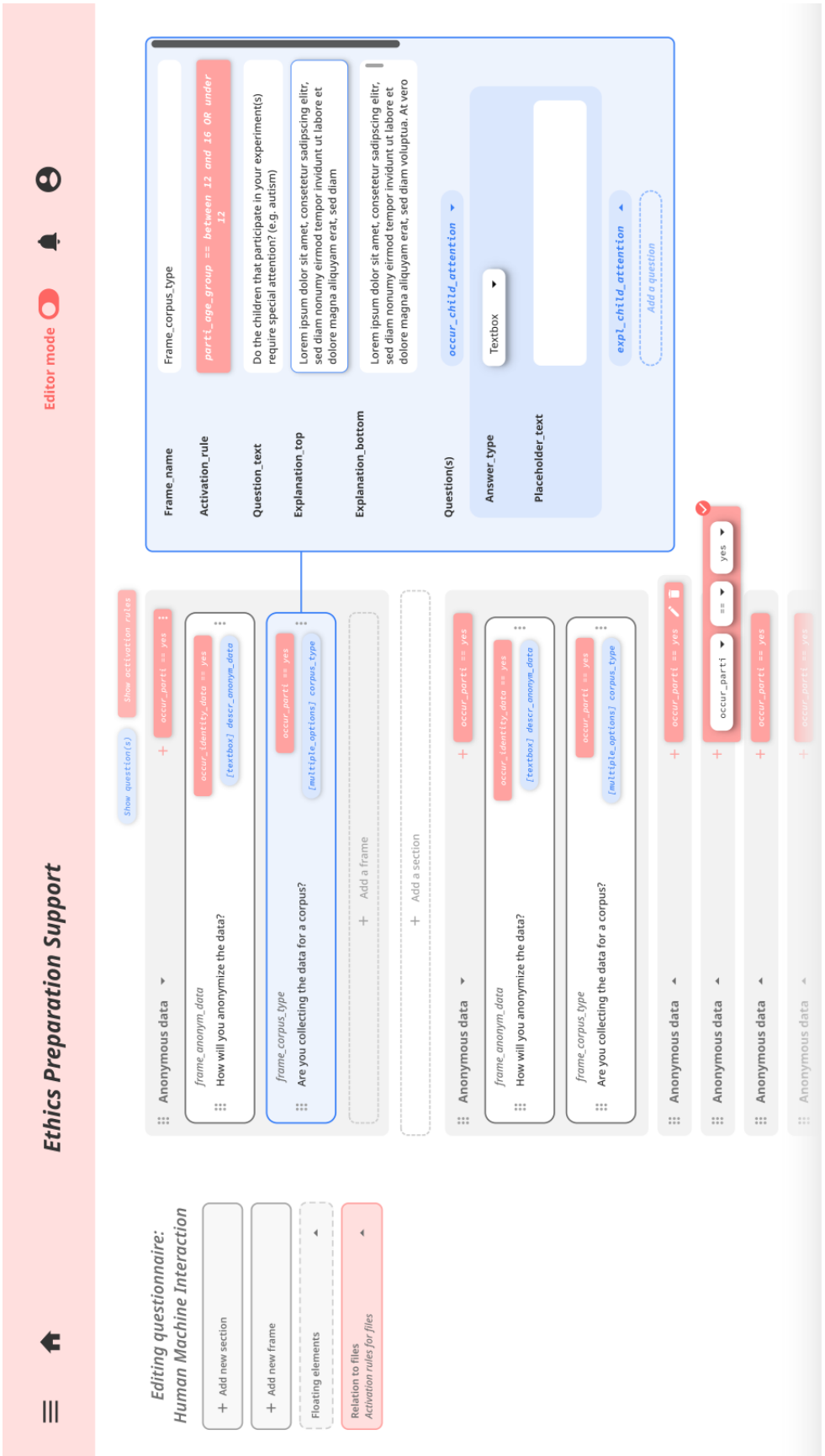


Figure 20: full overview design editor environment

F List of typical research

F.1 Types of research

- Observation of real-life activities
  Make sure you know how you deal with accidental inclusion of non-participants Is there any effect of the research on the safety of the activity, does it disturb the normal activity
- Interaction with a prototype
  Is there any danger in adverse effect during this activity (injury, shock etc.)
- Co-design session
- Deceptive research
  - Why is it absolutely necessary for your research goals
  - Are participants being deceived about risks
  - Is the participant debriefed after research
  - Is retroactive withdrawal of consent possible
- Interviewing experts on their expertise
  - They are not the subject of research
  - They give written informed consent
  - Adhere to current data management standards

F.2 Different participants

In this case, everything that has to be considered for a normal participant also has to be considered for the vulnerable participant. for everything: inclusiveness

- Research with non-vulnerable participants
  They should give informed consent before participating and have knowledge about risks/benefits of study
- Research with children participants
  Consent given by parents caregivers. Depending on their age it can be desirable to provide the children with some form of information/consent. Does your language fit the age group. A teacher/caregiver present at experiment can be desirable depending on activity and age of children to protect their interest. Stop protocol needs to be done more carefully (relates to voluntary / possible to withdraw → bepaalde secties in checklist)
- Research at institutions or schools
  Schools and institutions might have their own regulations to adhere to people in the location might not expect that research is going on
- Research with elderly participants
  Provide them with information in a way they can understand. Think about language, difficulty reading small fonts. Keep their generally lower proficiency with technology in mind alzheimer / need for proxy risks? (if so check that box!)
- Research with people with physical disabilities/illness
  Be mindful of signals of pain and trauma during experiment. Depending on the illness or disability, discuss experiment with doctor/expert
- Research with people with mental disabilities
  Is person able to give informed consent? If not who is responsible for this stop protocol: Be mindful of signals of pain and trauma during experiment. Possibly Involve caregiver. When in doubt, discuss experiment with doctor/expert.

- Research with people with addiction/mental illness
  Is person able to give informed consent? If not who is responsible for this Be mindful of signals of pain and trauma during experiment. Possibly Involve expert/doctor. When in doubt, discuss experiment with doctor/expert.
- Research with participants from specific ethnic,cultural groups
  Think about if there are any factors that could complicate the consent process and experiment (language, not wanting to say no, taboo topics etc.).
- Research with participants who are low literate or speak little Dutch and English
  Present all information (written or orally) in a way that fits the level of understanding of the participant. possibly check understanding trough asking questions.
- Research on individuals dependent on researcher (power relation)
  The higher the risk of the research, the more problematic power relations are research with proxy users (students stand in for people in rehab; colleagues stand in for ppl with mental disab; etc) → risks of representativeness.

F.3 Activities during research

- Having participant do physical activity
  Consider risks of the activity itself. Do they change because of the experiment Have participants interact with each other.
- Keeping a diary with personal information
  Clarify for all datapoints why you track them. Think about how to anonymize well enough that the participant cannot be de-anonymized when data is combined
- Using VR
  Vr-sickness, make sure you check for this specifically and know what to do when people get sick.
- Using wearable
  Consider the comfort of the participant, of both genders, during the time of wearing.
- Interacting with digital prototype
  Is all personal information entered stored securely.
- Interacting with physical prototype
  Is it safe to use.
- Recording emotional states
  What are possible biases of the way you record emotional data.
- Filling in questionnaire

- Watching/ viewing audio video material
  Can the contents be upsetting to some participants.
- Having participant eat or drink something
  Possible allergies
- Using or developing AI technology
  Are there possible biases in your training data, how do you mitigate harmful biases. Will your developed AI possibly influence human behaviour in an unexplainable manner. Are there possible misapplications or dual uses for your AI. Will it be possible that the AI violated someone's privacy
- Possible cybersecurity issues
  Chance of accidental discoveries of vulnerabilities. If you use malicious software, make sure you identify and mitigate risks.
- Research on social media, forums, etc. (also see entry under "collecting personal data")
  Are creators of data the subject themselves or is the data the subject of research.

- Being measured in some way in daily life (via phone)
  How do you ensure only the data you need for research is tracked. How do you end tracking after research. Clarify for all data points why you track them. Think about how to anonymize well enough that the participant cannot be de-anonymized.
- Interact with AI
  Is there a possibility of misinterpreting the AI decisions. Is it possible that participants mistake the AI for human.

F.4 Mode of collecting personal data

For all of these :

Is the data you collect necessary for your research? Think for each type of data what the goal of collection is. Is data stored according to the current regulations for storing personal data. Until what point in your research do you need to keep the data Are people aware until what time/moment they can request deletion of data

specific situations:

- Collecting data via surveys
  Maybe something about what information do you actually track with your question but maybe thats out of the scope of the tool.
- Collecting audio data/Collecting video data
  Is there a possibility of accidentally capturing non participants
- Collecting physical data (heart rate etc.)
  is there a chance of finding an anomaly? If so how do you notify participants (maybe consult the METC)
- Collecting observational data
  Think about the observation protocol/coding. Can you record data objectively.
- Collecting device sensor data (phone)
  Is the other information on the device still secure? is this influenced in any way though the collection of data.
- Collecting basic personal information (name, age etc.)

- Collecting data from diary study
- Collecting semi-public data (social media etc.)
  Is your data public, semi-public or private. Is it likely creators of data are aware of the possible use of their data for research.
- Collecting data using EEG/ECG
  is there a chance of finding an anomaly? If so how do you notify participants (maybe consult the METC). possible allergic reaction on electrode gells. Think about what you should do if that happens.
- Collecting data using wearables
  Can the setup be uncomfortable for the participant?
- Using existing data sets
  Is data gathered explicitly for scientific research and does your research fall within the scope of what the data was originally collected for If from public sources and does the source allow for using the data in their therms and conditions.

G Materials evaluation

G.1 Case description

Description of research proposal

- **Goal of the study:** evaluating a new prototype for a puzzle game for children between the ages 8-12
- **Goal of the puzzle game:** making doing simple math problems more fun for children
- **Participants:** 20 children between the ages 9-10 of a local school. Teachers have not mentioned any special disabilities.
- **Who will be present:** Two researchers conducting the experiment, one teacher and one assistant from the school to comfort, keep an eye on and manage the children
- **Location of the study:** Lecture room in the design lab
- **Data used in the study:** written notes made by three researchers observing the children when they play the game, answers of a small interview, data from an evaluation of a previous prototype
- **Data will be used to:** compare the effectiveness and likability of this prototype and the previous version of the prototype. Data will of course be anonymized
- **Information about the previous study**: According to the documentation of the previous experiment participants in this experiment have given consent to the use of anonymised and aggregated data for further research towards developing an educational game for children.

Description of the new prototype: The new prototype consists of several cubes which display numbers on all sides. Children play together in a group and all get a cube. The base station (in this case a researcher) asks them to solve a simple math problem like “add and/or multiply numbers to make the number 88” which they can solve by combining the numbers of their cubes in a certain way.. The children have to discuss what the solution might be and create it by putting several cubes in a certain pattern or order. If the solution is correct the base station(researcher) will compliment them.

Test setup and close-up of boxes

G.2 Demographic questions

- What year of your study are you in?
- Why did you decide to participate in this study? What is your motivation?
- Do you have experience doing research with participants?
- What is your experience with ethics assessment if you have any?

G.3 Questionnaire

all negative questions are marked in red.

Motivation

- The task of preparing for the ethics assessment was... [Difficult to perform]
- The task of preparing for the ethics assessment was... [Not interesting]
- The task of preparing for the ethics assessment was... [Exiting]
- The task of preparing for the ethics assessment was... [Easy to perform]
- The task of preparing for the ethics assessment was... [Clear]
- The task of preparing for the ethics assessment was... [Boring]
- The task of preparing for the ethics assessment was... [Confusing]
- The task of preparing for the ethics assessment was... [Interesting]

Dependability

- The advice and other documents provided by the tool were... [Unpredictable]
- The advice and other documents provided by the tool were... [Obstructive]
- The advice and other documents provided by the tool were... [Does not meet expectations]
- The advice and other documents provided by the tool were... [Supportive]
- The advice and other documents provided by the tool were... [Predictable]
- The advice and other documents provided by the tool were... [Meets expectations]

Effectiveness

- The advice and other documents provided by the tool made process of completing the ethics assessment... [Inpractical]
- The advice and other documents provided by the tool made process of completing the ethics assessment... [Fast]
- The advice and other documents provided by the tool made process of completing the ethics assessment... [Organised]
- The advice and other documents provided by the tool made process of completing the ethics assessment... [Slow]
- The advice and other documents provided by the tool made process of completing the ethics assessment... [Inefficient]
- The advice and other documents provided by the tool made process of completing the ethics assessment... [Messy]
- The advice and other documents provided by the tool made process of completing the ethics assessment... [Efficient]
- The advice and other documents provided by the tool made process of completing the ethics assessment... [Practical]

Self-competence

- I feel confident in my ability to do an ethics assessment without assistance. I am capable of learning how to fill in an ethics assessment in the future.
- I am able to achieve the goal of getting my ethics assessment approved in a future project.
- I feel confident of being able to deal with challenges that could occur during the ethics assessment process.

Perceived choice

- During the task I always felt like I choose the things I do.
- During the task I did what I was told to do, but I don't feel like I had much input.
- I do not feel free to complete the task in a way that I think is right.
- I did what I have to to complete the assignment, but it did not feel like I had any choice about how I did it.

G.4 Interview questions

- What do you think about the way the advice was given? What type of advice do you think is most useful to you?
- Was there anything unclear to you? if yes, what and how could it be improved?
- What did you miss in the advice given?
- What part did you think was very useful?

H Additional results

Group	Motivation	Dependability	Effectiveness	Self-comp.	Perceived_ch.	Correct Q.
1	3,5	5,3	3,7	4,7	4	17
1	3,2	5,5	5,6	5,2	6,5	8
1	5,2	5,8	5,4	5,5	6,2	15
1	6,4	6,5	6	6,5	6,5	14
1	5,6	6	7	6,7	6,5	8
1	4,6	5,7	5	5,7	6	17
1	5,9	6,3	6,2	4,2	5,2	10
1	3,7	5,3	5	4,7	5	14
2	5,1	5	5,5	6,7	7	16
2	4,5	4,5	4,9	5	4,7	17
2	4,2	5,3	6	5	6,2	9
2	3,5	5,5	5,5	6,5	4,5	15
2	5,2	6,2	6,1	6	6,2	16
2	6	6,2	6	5,7	6,2	12
2	4,5	5,3	5,1	6,2	5,2	16
2	5,7	5,8	5,7	6,2	6,5	10
2	4,7	6	5,6	4,7	5,5	9
3	4	4,3	5	5	4,2	11
3	3,4	4,3	5,1	5	4,2	5
3	5,5	6	6,2	4,7	5,2	4
3	3,9	5,3	6	6	4,5	17
3	3,2	4,7	3,6	3,5	5,2	8
3	4,5	4,8	5,6	3,5	4,5	6
3	1,9	2,5	2,5	3,7	6	7
3	4,5	4,7	4,9	4,5	5	11

Table 4: Averages of all metrics and correctly answered questions per participant rounded to one decimal.

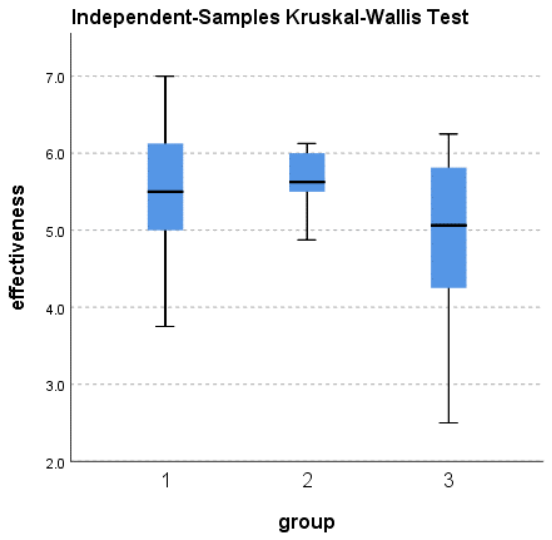


Figure 21: Boxplot effectiveness

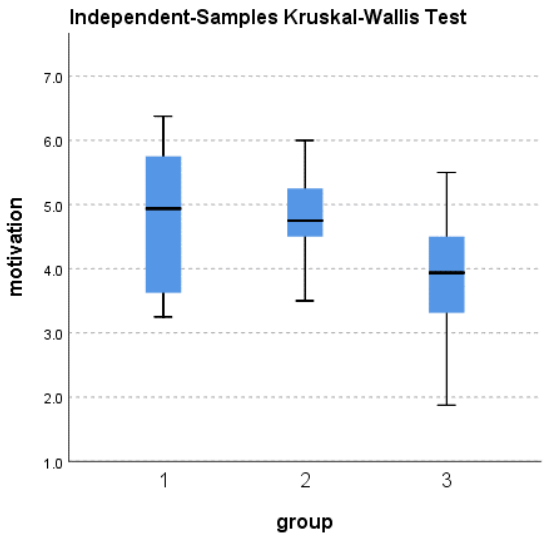


Figure 22: Boxplot motivation