

The effect of an educational escape room on master student motivation towards research integrity

Research project

Author: Marco Verhoeven

Student number: 5512379

ECTS: 30

Utrecht University

Graduate School of Teaching (GST)

Supervisor: Dr. R.P. (Roald) Verhoeff

Second examiner: Drs. A. Veldkamp

Faculty: Freudenthal institute

Date: 01-07-2021

Table of Contents

1.	Abstract	4
2.	Introduction.....	4
3	Theoretical framework.....	6
3.1	Research integrity (RI) and responsible conduct of research (RCR).....	6
3.2	Empowerment.....	7
3.3	Motivation	8
3.4	Gamification and game-based learning.....	9
3.5	Educational escape rooms.....	10
3.5.1	Escape room puzzle structures.....	12
4	Methods	13
4.1	Research approach/strategy	13
4.2	Target group/Participants	14
4.3	Online platform	14
4.4	Activity procedure	15
4.5	Data collection.....	15
5	Design process.....	17
5.1	Design pre-pilot	17
5.1.1	Rooms.....	19
5.1.2	Routes.....	21
5.1.3	Dialogues and grey dilemmas.....	22
5.1.4	Storyline/Narrative.....	25
6.	Pilot.....	26
6.1	Pilot evaluation.....	26
6.2	Escape room redesign	27
7	Results	29
7.2	Interest/Enjoyment	29
7.2	Perceived competence	33
7.3	Effort/Importance	38
7.4	Pressure/Tension.....	42
7.5	Perceived choice.....	44
7.6	Relatedness	46
7.7	Value/Usefulness.....	48
8	Discussion.....	51
8.1	Qualitative result interpretations.....	51
8.2	Study approach and methods	53

9	Conclusion	54
10	Further research	56
11	Re-evaluation and design recommendations.....	57
12	Acknowledgements	59
13	References	59
14	Appendices	63
14.1	Appendix A: Observation form.....	63
14.2	Appendix B: Observation example	64
14.3	Appendix C: Post-activity survey	65
14.4	Appendix D: Focus group instrument.....	66
14.5	Appendix E: Narrative/Storyline practical elaboration	67

1. Abstract

Research integrity (RI) is an important aspect of conducting research for (future) researchers, especially recognizing and dealing with unforeseen situations such as grey dilemmas. To increase master student motivation towards research integrity, an engaging educational tool was needed. An educational tool with potential, is the educational escape room (ER). Globally, educational escape rooms became increasingly popular. This study designed an educational escape room activity about research integrity. The design focuses mainly on motivational aspects, which are important to eventually empower master students. Several game-based learning and gamification elements are used in the design to generate engagement and motivation towards research integrity. The design was evaluated by running a pilot and three test groups, consisting of four people per group. Most participants thought the escape room was fun, engaging and challenging. Time pressure caused extra effort, but less elaborate discussions and the design context proved to be meaningful and recognizable to several participants. Opposite results showed a lack of direction inside the rooms, which caused frustration. Participants also felt distant towards group members because of the online setting. The results of the study suggest that the master students felt more engaged and motivated towards the completion of the escape room activity, but the escape room was not able to create an optimal environment where all the participants' needs for intrinsic motivation were fully saturated. The rather positive results still confirm the potential that the educational escape room activity has to affect master student engagement and intrinsic motivation towards research integrity positively.

Key words and concepts: Educational escape room (activity), design study, research integrity (RI), grey dilemmas, game-based learning (GBL), gamification, (intrinsic) motivation, engagement, motivational elements, master students

2. Introduction

For academic research it is of great importance to achieve and retain the public trust, else reliability of research results will constantly be questioned. To gain trust in scientific research, research is conducted as integer as possible.

Research integrity (RI) forms the basis for continued investment in research and reliance on scientific findings in decision-making (Anderson, Shaw, Steneck, Konkle and Katama, 2013). Research integrity and misconduct refer to *right* and *wrong* behavior in research. Policy makers and the research community have discussed and debated about how to label, study and respond to research behaviors that do not live up to the standards of responsible conduct for the last 25 years (Steneck, 2006). Eventually three categories within research integrity emerged, separating three different types of research behavior: responsible conduct of research (RCR) fabrication, falsification and plagiarism (FFP), questionable research practices (QRP). RCR resembles the ideal behavior in research practices, whereas FFP stand for the worst possible behavior in conducting research. QRP are practices in between the other two categories. These practices do not impact the integrity of the research process in a direct and negative way (Steneck, 2006). An example of a questionable research practice is publishing the same research results in two or more publications.

While conducting research, researchers could encounter unforeseen and complex situations and dilemmas. Some dilemmas are just solved with a *yes* or a *no* answer, but most dilemmas are more complex with several viable answers. An example is the free-rider problem, where one or more group members are not putting as much energy into group effort as other members, but still getting the same result (Kim & Walker, 1984). There are many feasible options to solve this problem, like discussing the matter with a supervisor or just leaving his/her name of the final product. The free-rider problem makes a fine example of a *grey dilemma* in research. Most people know for example that plagiarism is *bad* research behavior and that citing properly is *good* research behavior, but

conducting research responsibly goes combined with making well-thought choices and decisions when encountering issues like the free-rider problem. Globally, there is a need to increase the awareness of future researchers to avoid misconduct and (intentional) questionable research practices, but maybe more importantly, to prepare future researchers on how to conduct research responsibly and to learn them how to make responsible choices and decisions when encountering grey dilemmas like the free-rider problem (Turens, 2005; Robshaw, DeMets, Wood, Boiselle & Hennekens, 2020; Steneck, 2006).

An online study on the attitude of University students at two institutes in Europe and Africa towards research integrity show interesting outcomes. These institutes are the International Livestock Research institute (ILRI) and the Swedish University of Agricultural Sciences (Lindahl & Grace, 2018). Of 138 participating students and 14 participating supervisors, 98% of them had heard about plagiarism and 35% believed it was common. Only 45% had heard of plagiarism done by the author himself and 30% thought it was acceptable to cite a reference from a paper they had not read. These are some result examples from the study. These outcomes indicate that at these institutes, students are in need of clarification on certain aspects of research integrity like FFP and QRP. However, this study is about only two institutes from two different continents with different types of higher education. So the results might seem less relevant than when more institutes got involved, but the study still highlights the need for more student awareness on research integrity.

An addition to these results, are the survey outcomes of a study on the teaching of research integrity in Europe, as part of the H2020 INTEGRITY project. A team consisting of Andorno, Katsarov & Rossi (2020) conducted the survey. The H2020 INTEGRITY project is a collaboration between 11 partner organizations, including the Ethics Institute at Utrecht University. The survey outcomes indicate a few important points: 30% of the 98 research integrity teachers that participated to the survey suggest that they struggled with learners who question the relevance of a course about research integrity and lack personal motivation. According to respondents, the teaching of research integrity in their institutes is focusing too much on spectacular examples of misconduct (FFP). The respondents would rather learn about questionable research practices and unforeseen situations in the teaching contents. The respondents also spoke out importance of certain role models in conducting research, like supervisors and senior researchers. They would like to involve these supervisors and senior researchers in the teaching on research integrity, because the behavior of the role models in conducting research greatly influences the behavior of future researchers.

When looking at a local evaluation the students filled in on their master course at the UU, one course stands out from the rest. A course on research integrity scores consistently quite low (3.0 on a 5.0 point scale) and the master students lack interest in learning about research integrity, developing their skills in conducting responsible research. The lack of interest that master students have towards research integrity, suggests that a more innovative, motivating and engaging approach with appealing educational tools is needed for the research master students (Clarke, Arnab, Keegan, Morini, and Wood, 2016; Clarke, Peel, Arnab, Morini, Keegan & Wood, 2017). A promising tool that implements these specific elements, is the educational escape room (ER).

Escape rooms are collaborative, puzzle-solving games and are popular worldwide (Nicholson, 2015). Since a couple of years, escape rooms are used for educational purposes to motivate students and to make learning about a subject more fun and engaging (Clarke et al., 2017). An example is the rather negative attitude several students have towards mathematics in secondary education. Most students think that lessons on mathematics and developing mathematical skills are not very interesting, but still important (Kislenko, Grevholm & Lepik, 2007; Hoskonen, 2007). Escape rooms are being used to make learning about math and other complex or dry subjects more fun for the students to get them motivated, engaged and possibly even changing their attitude towards the

subject (Glavas & Stascik, 2017). Therefore, an escape room activity on research integrity has great potential to stimulate research master student motivation, engagement and interest towards research integrity as well.

The escape room activity designed in this study, aims to provide an engaging and motivating experience for research master students towards research integrity. Besides motivating the master students, empowering them is the starting point of the escape room. Empowerment in general defines as the students' strengths, competencies, natural helping systems (everyday life relationships with family members, friends etc.) and proactive behaviors to social policy and social change. The reason for this type of starting point for the escape room is of its positive nature. As indicated by the survey results of Adorno, Katsarov & Rossi (2020), teaching on research integrity in European institutes focus too much on bad research practices (FFP). This escape room offers the straight opposite, based on how research should be done (RCR) rather than should not be done (FFP). QRP are also addressed in the escape room, as answer options to grey dilemmas. This information generated the following research question and sub-questions:

Research question

How does an educational escape room on research integrity affect the motivation of master students towards research integrity?

Sub-questions

1. How are master students motivated during the escape room activity on RCR?
2. How do escape room elements affect the intrinsic motivation of master students?

3 Theoretical framework

3.1 Research integrity (RI) and responsible conduct of research (RCR)

"Research integrity is about upholding the highest standards when engaging in research, not merely avoiding wrongdoing or questionable research practices" (Nichols-Casebolt, 2012, p. 4). When using the term *integrity*, some words that come to mind are truthfulness, accuracy, honor, rightness and honesty (Anderson et al., 2013). Research integrity stands for doing research and analyzing research results the *right* way. Research integrity applies to multiple stages of conducting research and one important stage is analyzing and interpreting research findings. Research findings are communicated to the outside world. Most people take these findings for granted because they assume that the research was conducted responsibly. This accountable behavior of doing research is called responsible conduct of research (RCR). "RCR is simply *conducting research in ways that fulfill the professional responsibilities of researchers, as defined by their professional organizations, the institutions for which they work and, when relevant, the government and public.*" (Steneck, 2006, p. 55). Responsible conduct of research is a part of research integrity as a whole, but what does it mean to conduct research responsibly? At a minimum, the researcher will not falsify data, steal work from other researchers and stands in for the protection of participants of the research: "In general terms, responsible conduct in research is simply good citizenship applied to professional life. Researchers who report their work honestly, accurately, efficiently, and objectively are on the right road when it comes to responsible conduct" (Steneck, 2007, p. 11).

The opposite of RCR is research misconduct in the forms of fabrication, falsification and plagiarism (FFP) (Anderson et al., 2013, p. 219; Steneck, 2006). Fabrication is the intentional misrepresentation of research results, falsification is the manipulation of research results and plagiarism is using and presenting work or ideas from someone else as your own, with or without their consent. These are the most serious forms of misbehavior possible in research. QRP are also

seen as misconduct when done intentionally. The difference between FFP and QRP is that QRP are seen as less serious and are not damaging the integrity of the research process directly (Steneck, 2006). Well-known examples of QRP are result misrepresentation, inaccuracy and bias. Misrepresentation in research means that researchers are not honestly and accurately presenting their contributions to research publications. Inaccuracy in research stands for the exclusion of honest errors and careless mistakes in the research process by researchers. When researchers are not making efforts to separate personal, subjective views from experimentally based factual information, this is called bias (Steneck, 2006). Besides the type of behavior researchers are able to have while conducting research, other complex and unforeseen situations can occur during the research process. Such complex issues that master students are rather unaware of, are called grey dilemmas.

But what are these grey dilemmas in research exactly? Grey *dilemmas* should not be confused with grey *areas* in research integrity. Grey areas in research integrity are the areas in between the black- and white areas. The black areas represent the 'wrong' areas like FFP and QRP when done intentionally, whereas the white areas are the good areas like RCR. Grey areas "are conceptually open to include any potential breach of integrity, thereby embracing the "grey areas" of research conduct" (Fanelli, 2011, p. 83). These areas are mostly QRP when done unintentionally. An example of a grey area in research is the neglect of negative outcomes. Grey areas in research integrity are not the main focus of this design study, but are still used in the escape room design. However, grey *dilemmas* are more important aspects used in the escape room design. The introduction states a recognizable example of a grey dilemma, the free-rider problem. Many students come across this type of dilemma in their own research practices. What to do when a group member decides not to contribute to the work (he or she just rides along with the work of the others)? When discussed in a group, multiple viable answers will pop up. Some of these options are better than others, but even the lesser options could still be viable. A lot of master students are unaware of the fact that they are able to encounter grey dilemmas while conducting research and often do not know how to recognize them and deal with them confidently.

3.2 Empowerment

The Escape Room project aims to empower the students, so starting at their strengths, competencies, natural helping systems and proactive behaviors to social policy and social change (Rappaport, 1981; 1984). There are three dimensions of student empowerment: academic, political and social empowerment. James Cummins (1986) stated the following regarding academic empowerment: "Develop the ability, confidence, and motivation to succeed academically. They participate competently in instruction as a result of having developed ... appropriate school-based knowledge and interactional structures" (p. 22).

As stated by James Cummins, one of the reasons why empowerment is important is because it develops skill, confidence and motivation to succeed in education. A lack of empowerment could cause a lack of student engagement, confidence and motivation in learning (McQuillan, 2005). Empowered students also internalize "higher level cognitive skills and assume greater control over setting their own learning goals" (McQuillan, 2005; James Cummins, 1986, p. 28). Empowering research master students will increase their engagement towards research integrity education and will grant them higher cognitive skills and greater control in their own learning process and setting up learning goals. To empower master students towards research integrity is a rather innovative approach. The aim of the approach is to increase student awareness on unforeseen dilemmas in research, like grey dilemmas. The need for a more positive and empowered approach is high, because research integrity education in general focusses more on preventing misconduct than on

learning how to conduct research responsibly. Van den Hoven & Krom (2020) stated: "We should empower (future) researchers instead of floating them in a corner of misconduct".

To achieve student empowerment, several cognitions and needs that students have are important to integrate in educational tools. These cognitions originate from empirical models based on employees, rather than on students (Arciniega & Menon, 2013). However, the cognitions are useful for students as well. According to Menon (2001, p. 161) an empowered state is a cognitive condition "characterized by perceived control, competence and goal internalization" (Arciniega & Menon, 2013, p. 2950). Perceived control is the first cognition and defines as a need that is fulfilled when a student is in control of their actions and their environment at all times. The second cognition is competence or perceived competence. When a student feels competent, he or she has feelings of self-efficacy. Self-efficacy stands for "an individual's belief in his or her capacity to execute behaviors necessary to produce specific performance attainments" (Bandura, 1977, 1986, 1997). Goal internalization is the last defined empowerment cognition, which is achieved when a student understands why the activity he or she has to do, is relevant or significant. To integrate these cognitions into an escape room activity, it is important that the students have control over the situation they are in, that answers they give and actions they perform are not necessarily labeled as bad or wrong and that the context and contents of the escape room are meaningful and valuable to the students. But before a student is able to become empowered, the motivation, engagement and interest towards research integrity are factors (among others) that need to be addressed first.

3.3 Motivation

When using the term motivation, the words *must* and *want* are of important note. Someone can do something because he wants to do it or because he must do it. These are two different motivators for people and these motivators are also observable in education. When someone must learn because of external stimulation, he or she is extrinsically motivated. When someone is extrinsically motivated, they are motivated to do something because of external factors like rewards or grades. When someone wants to learn because of internal stimulation, there should be a feeling of autonomy, competence and relation towards the act or behavior. These three needs of the human being are part of the self-determination theory, created by Ryan and Deci (1985; 2010; 2017). The self-determination theory compares intrinsic and extrinsic motives to focus on the degree at which the behavior of an individual is self-determined (and self-motivated). When all the three needs are satisfied, someone should be intrinsically motivated (self-motivated): being naturally motivated to do something, out of interest rather than for reward. *Autonomy* stands for freedom of choice towards own insights and has the possibility to influence own actions. *Competence* resembles the trust that someone should have in own capabilities and *relatedness* is the feeling of social connection and trust in the environment. The self-determination theory is also used to interpret motivational situations regarding students in educational settings. When teachers increase the feelings of the students towards the three basic needs, the overall intrinsic motivation of the students should increase.

Intrinsic motivation proved to be a good indicator for student performance (Cerasoli, 2014). This indicates, that an intrinsically motivated student would perform better than a student who is not. A review study on predictors of academic performance and school engagement shows that higher academic performance was obtained by students via higher elevation of motivational and persistence factors, which proves that motivational models are strong predictors of both secondary students' and higher academic performance (Moreira, Dias, Vaz & Vaz, 2013). Also, students become often less motivated as they progress their school career when they are extrinsically motivated (Otis, Grouzet & Pelletier, 2005) and intrinsic motivation shows better school outcomes than extrinsic motivation (Ryan & Deci, 2008).

These statements based on empirical studies, suggest that extrinsic motivation is a *bad motivator* for students. However, this is not necessarily true in every situation. Not only is extrinsic motivation more present in educational settings than intrinsic motivation (Harpine, 2015), it is also a very good motivator for people who have little interest in performing an activity or learning about something they just do not care about (Tranquillo & Stecker, 2016). It is also less difficult for a teacher to achieve extrinsic student motivation compared to intrinsic student motivation, with grades providing the most popular educational extrinsic motivator.

3.4 Gamification and game-based learning

It is rather difficult to increase (intrinsic) student motivation towards certain subjects like research integrity. Gamification and game-based learning are methods that both aim towards making learning more fun and engaging for students. Gamification defines as the integration of game-elements and game-mechanisms in non-gaming environments (Deterding, Dixon, Khaled, & Nacke, 2011; Kapp, 2012; Huang, & Soman, 2013). When used in education it is called *educational gamification*. Besides the application of game-elements and mechanisms in education, there are also full games developed for educational purposes. The use of actual games in the learning process of the students is called *game-based learning* (Wendel, Gutjahr, Göbel, & Steinmetz, 2013; Al-Azawi, Al-Faliti, & Al-Blushi, 2016). The games used for educational and other purposes besides entertainment, are also called *serious games* (Michael & Chen, 2006). Gamification and game-based learning are both strategies that are able to increase student motivation and have a positive impact on student learning (Sawyer, & Smith, 2008; Buckley, & Doyle, 2014; Su, & Cheng, 2015; Papastergiou, 2009). Gamification consists of certain elements that are important for the effect and learning outcome of the student, such as: creativity, research, discovery, difficulty, competition, danger, feeling locked up, sensation, fantasy, sympathy, simulation, cooperation and disobeying rules (Costello, & Edmonds, 2007). The integration of these elements can determine the experience and learning performance of the student during an activity (Cerasoli, 2014; Moreira et al., 2013).

When using serious games in education, the design of the game is of great importance (Nacke, Drachen, Kuikkaniemi, Niesenhaus, Korhonen, van den Hoogen, Poels, IJsselsteijn, & de Kort, 2009). The goal of a successful serious game design is to stimulate *meaningful play*, according to Salen, Tekinbaş and Zimmerman (2004) (see figure 1). Meaningful play occurs when certain game design characteristics are recognized by the participants, such as being autonomous, the ability to come up with short-term and long-term strategies, the ability to react on changing situations and being able to absorb- and process information to accomplish tasks as fast as possible (Van Bree, 2011; 2013). Also, good instruction and direct feedback at the end of the game are important factors when designing a serious game. Meaningful play is important for reaching motivational and empowering goals regarding master students. When integrated efficiently, students should feel that the escape room has value to them. When the escape room contents are valuable and meaningful to the students, it is possible that their goal internalization (one of the three empowerment cognitions) and motivation towards the escape room and eventually towards research integrity in general will be influenced positively.

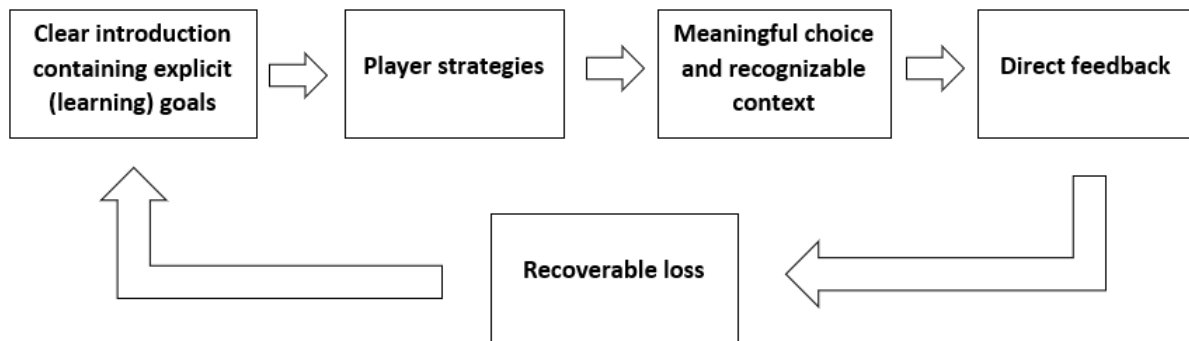


Figure 1. Meaningful play model, adapted from van Bree (2013)

Another principle in games in general, is recoverable loss. When playing a lot of games, the progression of the player gets lost after losing to a part of the game. Recoverable loss stands for the recover capacity of player progression while playing the game and failing to complete a part of it. Escape rooms are mostly lacking recoverable loss, because the contents and solutions are already known to the participants from the first time, when they compete for a second time.

The last important factor to take into account when using serious games, is the mental state of the participants. Csíkszentmihályi (1990) created the *flow theory*: a student is in a ‘flow state’ when the student is fully engaged in the activity, when the activity connects with the capacity of the student and when the goals of the activity are achievable and clear with direct feedback at the end of the activity. A student that is in a flow state during the activity, functions optimally (Csíkszentmihályi, 1990; Nakamura & Csíkszentmihályi, 2002; Csíkszentmihályi, & Larson, 2014; Pavlas, Heyne, Bedwell, Lazzara, & Salas, 2010).

3.5 Educational escape rooms

An educational escape room is categorized as a serious game and is a rather innovative educational tool. “Escape rooms are live-action team-based games where players discover clues, solve puzzles, and accomplish tasks in one or more rooms in order to accomplish a specific goal (usually escaping from the room) in a limited amount of time” (Nicholson, 2015, p. 1). The global success of the escape room comes from all kinds specific game elements and are becoming more popular as a tool to engage students in their educational setting (Kinio, Dufresne, Brandys & Jetty, 2019; Wiemker, Elumir, & Clare, 2015). A first example of a game element present in escape rooms is the social element, consisting of cooperation and teamwork. Discussing and working together to solve puzzles or riddles is fun and this makes escape rooms ideal for groups of people. A second example is the competition element, which motivates participants extrinsically to beat other groups. It triggers adrenaline and sensation which causes people to enjoy themselves while participating in the escape room. A third example is that escape rooms have endless options for themes, storylines and fitting puzzles, riddles and tasks. These are elements which caused the education to create interest in integrating the escape room to educational settings, but not all elements are equally effective for educational purposes.

Adapting the elements or using other elements is necessary for the implementation of escape rooms in education. Elements that are important for educational escape rooms are: cooperation/teamwork, creatively finding solutions and critical thinking (Wiemker et al., 2015). To implement cooperation and teamwork in an escape room on RCR, dialogues and grey dilemmas are created. Students need to provide answers to the dilemmas to progress through the rooms. They have to work together, discuss opinions and think critically in order to come up with a well-thought

answer. The escape room further consists of puzzles and a time limit per room. Therefore, it is not possible to complete a room without good communication and teamwork. The puzzles also have research contexts that are recognizable for students. The puzzles are increasing in difficulty, the further the participants progress in the activity. Two different routes of puzzles, with one more integer route than the other, stimulates critical thinking. Furthermore, escape rooms have so many variation possibilities that almost every subject in education is able to use an escape room tool. Lastly, a cooling-down period is necessary after the escape room where all the answers, hints and puzzles are being discussed plenary. This phase is needed for direct feedback and to complete the activity (Nicholson, 2015). The cooling-down period is a very important part of the participants' learning process. By discussing the participants' decisions and progress of the escape room, the participants will receive direct feedback and will understand the mistakes they made and learn from them. The cooling-down period will also provide direct feedback and confirmation of choices they made during the activity. This will affect their feeling of competence.

Educational escape rooms have the potential to increase student motivation and also some research results suggest that motivation and engagement are increased by the use of the escape room concept in educational settings. The first example is a 'Room Escape' teaching experience, designed to improve student motivation and their learning towards two courses on a high education computer engineering degree (Borrego, Fernández, Blanes & Robles, 2017). The activity was divided into two phases. Students must first earn access to the escape room activity by solving challenges and tasks given during regular classes. The first students solving each challenge and completing each task get access to the escape room activity. During the activity, the students are in a locked room while solving puzzles, challenges and tasks while working together to get out of the room and winning the activity. The results of the escape room activity were rather positive: students participated actively in the escape room access challenges which increased their will and motivation to learn about computer engineering. Also during the escape room activity itself, students thought the activity was 'challenging, demanding, interesting and amusing' (Borrego et al., 2017). Besides, learning by the use of games increases middle school student motivation and creates positive attitudes towards different subjects like mathematics, which also is proven with the use of an escape room game (Glavas, & Stascik, 2017). The third example is a nursing escape room used as a teaching game, which not only motivates the students but also allows the students to recall and apply the knowledge gained in class (Gómez-Urquiza, Gómez-Salgado, Albendín-García, Correa-Rodríguez, González-Jiménez & Cañadas-De la Fuente, 2019).

Clarke et al. (2017) created an educational escape room framework *the escapED framework*. The framework consists of six phases: determining participants, objectives, theme, puzzles, equipment and evaluation of the educational escape room. When using the framework as an approach in designing an escape room activity, the design will provide a more contextually immersive learning experience, which can be used and adapted in all kinds of learning contexts. The design and research process are discussed in chapter 4 and 5. Also, a specified research framework was used that was adapted from the educational escape room framework that Clarke et al. (2017) created. The framework is also based partially on educational escape room design guidelines, created by Eukel & Morrell (2020).

The discussed empirical examples and results suggest that escape rooms made for educational purposes have the potential to increase student motivation and performance, and the teaching method proves suitable for multiple target groups varying from secondary education content to higher educational subjects.

3.5.1 Escape room puzzle structures

Puzzles within an escape room activity could be ordered following certain puzzle structures. Nicholson (2015) showcased four types of puzzle structures: *open*, *sequential*, *path-based* and *pyramid* (see figure 2 for a visual adaptation of the structures). The circles represent the puzzles and the rectangles represent the final code for that room or other victorious circumstances (meta-puzzles). The rectangle does not need to be the end for the whole activity. When using four rooms for example, the rectangle only represents one final code for one room. The open, sequential and path-based organizations are basic puzzle structures, whereas the pyramid organization is a rather more complex type of puzzle structure. A pyramid structure is called a hybrid model, which starts with several path-based puzzle structures. After completing these puzzles, the activity narrows down to one last path-based structure towards the final code. Most puzzles are more complex than described in the basic structures and do look more similar to the pyramid organization. Most common uses of the structures in escape room activities, are path-based. With a path-based puzzle structure, a team is shown multiple different puzzle paths at the same time. Each path leads on its own towards the final code (victory condition). This structure is favorable, because every participant could then work on a different pathway at the same time. Sequential structures are also pretty popular in escape rooms. Participants get one puzzle and solving the puzzle leads directly to an answer, which unlocks the next puzzle in that sequence (a more linear structure). Open structures are less popular, because it is hard to create a flow-based experience. Participants are able to solve multiple puzzles at the same time, which each provides a piece of the final code.

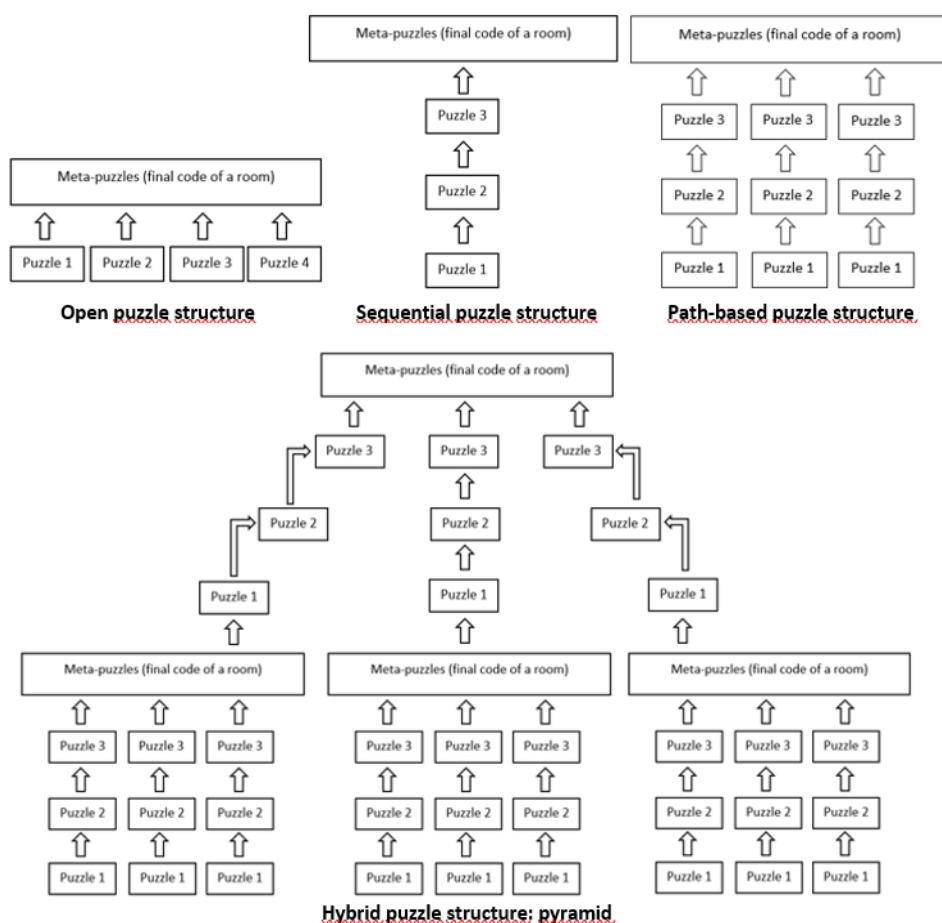


Figure 2. Escape room puzzle structures. Open, sequential and path-based organizations are 'basic', whereas a pyramid (hybrid) structure is 'complex'. The squares indicate the puzzles and the rectangles indicate meta-puzzles (final code for example). Adapted from Nicholson (2015)

4 Methods

4.1 Research approach/strategy

The type of research was qualitative with a design-based approach. The focus of the Escape Room project is to empower students by developing their skills for RCR, by increasing student awareness and helping students to reflect on (grey) integrity issues they will encounter in practice. The emphasis of the project is on the positive side of research integrity, so more on RCR and less on QRP and FFP. To gain this empowerment and skill development, the students need to be motivated and engaged towards research integrity content.

The escape room was designed and created together with master student Sarah Seghier. The focus of this study, differs from Sarah's perspective. This study had motivation and engagement as core variables, while Sarah focused mainly on empowerment and awareness. The collaboration went through all phases of the design study and was also present in the data collection (fellow observer and moderator) and data analysis (second coder). While using both perspectives, an escape room on research integrity was designed and created. The design contains elements that stimulate engagement and motivation and elements that stimulate empowerment and awareness.

The framework created for designing and studying an escape room on RCR, is based on the existing educational framework created by Clarke et al. (2017) and on educational escape room guidelines composed by Eukel & Morrell (2020) (See figure 3). The first phase of the process is designing and creating the escape room on RCR. This phase is further explained in the method section (chapter 4) and the designing process section (chapter 5). Phase two is designing and creating models for data collection, which is further explained in the method section (with the data collecting instruments observable in the appendices). Phase 3 describes the conducted pilot of the escape room on RCR on a group of master students. The pilot granted insights in improvement points and redesign criteria, which are further discussed in chapter 6. Phase 4 stands for the escape room test rounds that were conducted on three groups of four master students and the analysis of these results. The results (chapter 7), discussion (chapter 8) and conclusion (chapter 9) sections are providing more in-depth information on phase 4. The fifth and final phase of the framework is about the evaluation of factors that proved to have motivational potential and factors that are in need of improvement or replacement, regarding this particular escape room design. The re-evaluation and recommendation section (chapter 10) contains further information regarding phase 5.

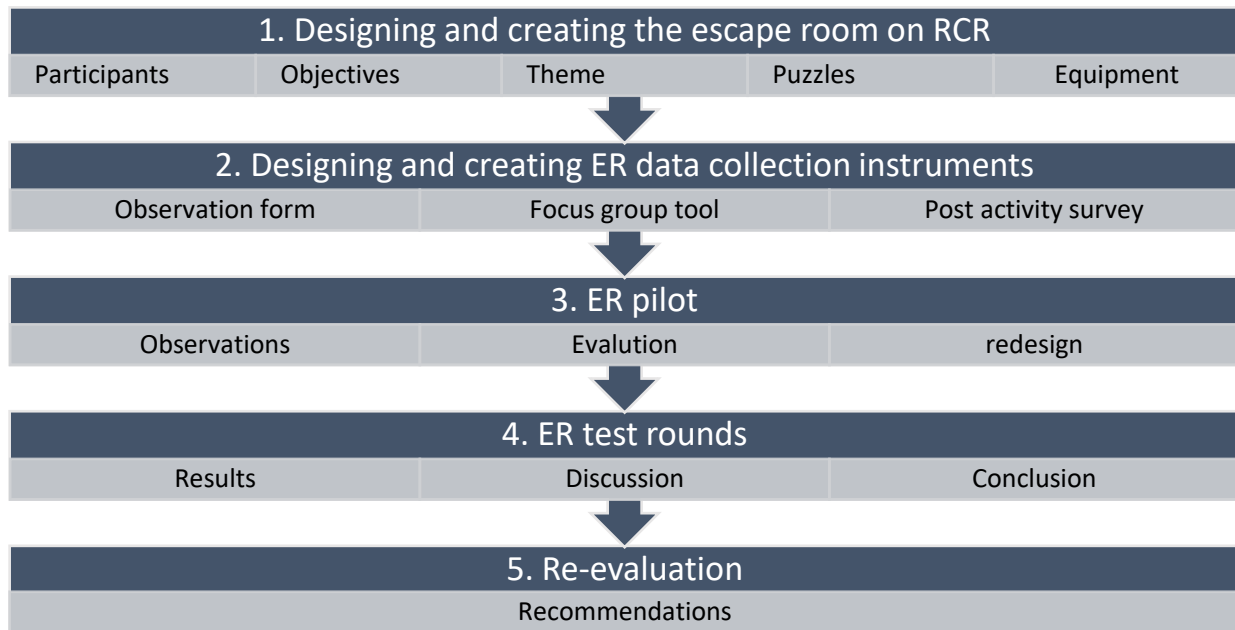


Figure 3. The used research process framework for designing, testing and evaluating this particular educational escape room activity on RCR. This framework is based on the educational escape room framework that is created by Clarke et al. (2017) and on the created guidelines to ensure educational escape room success by Eukel & Morrell (2020).

4.2 Target group/Participants

The participation aim for the escape room designed for this study was 6-8 groups of 4 people, which will be between 24 – 32 master students. The reason for this aim is because several studies suggest that anywhere from 5 – 50 participants for qualitative research is recommended as adequate (Dworkin, 2012). Five participants feels like too little because that would only be around one group of players. 50 participants would be too much, regarding the time limitation and corona virus circumstances. Around 24 – 32 participants is in between the 5 – 50 range and was the aimed sample size aim for the study. Unfortunately due to the circumstances, we have tested the escape room on only one pilot group and three test groups. The pilot group consisted of four participants and one spectating teacher, while the test groups all consisted of four participants each (16 participants in total).

The research master students that participated were specialized in all kinds subjects, like life sciences, natural sciences, humanities and ethics. The participants had an age ranging from 22 to 30 years old. The target group is approached via Roald Verhoeff and Mariëtte van den Hoven. An invitational letter was set up for potential participants from several university studies. Master coordinators promoted the invitational letter to their students. The student responses were then gathered and put in a date-picking system. A maximum of four people were picked for the available dates. These dates where scattered from December to March.

4.3 Online platform

The escape room activity was designed and created entirely online, in the desktop application of Microsoft PowerPoint. Although, optimal group game experience is gained via real-life elements, the choice of an online environment is in this case the most optimal way to test the escape room design (also due to the Covid-19 outbreak). The escape room, created in Microsoft PowerPoint is also easily

adoptable for any device that has the application installed. There are however a few disadvantages to the escape room created in PowerPoint.

Firstly, add-ins (like macros) that are used in the escape room are not always compatible for MacBook users. Windows devices usually have no problems with the escape room running in Microsoft PowerPoint with the add-ins. Secondly, the escape room file is very large. This causes older devices to run very slow or often freeze entirely, which is very frustrating for the participants. Thirdly, the escape room cannot be played properly in the online web version of Microsoft PowerPoint. The desktop application of Microsoft PowerPoint is needed to play the escape room activity properly. These disadvantages showed up before and during the tests performed in this study.

For this study during corona circumstances, an online platform was needed and Microsoft PowerPoint provided a platform that was *good enough*. The participants of the escape room activity participated in groups of four. These groups worked together online in Microsoft Teams (communicating) and PowerPoint (playing) on all different kinds of puzzles, tasks and challenges. These activities all consist of teamwork, but also a good division of tasks will be important.

4.4 Activity procedure

During the escape room the groups needed to escape the activity (by solving puzzles, completing tasks and discussing grey dilemmas) within the time limit, which was around sixty minutes in total (as almost all escape rooms are limited at sixty minutes). Right after the activity there was a cooling-down period. In the cooling-down period the moderators discussed most given answers to grey dilemmas, hints and puzzles. This part is essential for the escape room activity to complete it and grants direct feedback for the participants. According to the meaningful play model, this is very important for the learning process of the participants. Participants are able to ask questions and discuss answers with each other and the game moderator. For this study however, a complete debriefing on the escape room activity was not possible. The escape room was simply too big to completely discuss in 45 minutes. Besides, the participants got asked for two hours of their time. When discussing every aspect intensively, several hours would have been needed to completely discuss the escape room (and that is only the time that is needed for the cooling-down period).

Therefore, the cooling-down period was designed as a focus group. Remarkable decisions and remarkable notions made during the activity were discussed with the participants. A slight overview of the activity was discussed with the participants, but certainly not every aspect and every decision. For motivation, the questions of the focus group were based on the intrinsic motivational needs (autonomy, competence and relatedness). Also engagement, effort, value and pressure were motivational factors that were taken into account during the focus group.

A focus group question example: *Do you feel like you have learned more about how to deal with research integrity dilemmas through this experience and why or why not?* This question is focused on (perceived) competence and awareness (empowerment). However, when someone does not feel like he or she has learned something from the activity this might have been caused because other motivational factors lacked effect. When the relation (relatedness) between the group members is not good for example, the feeling of perceived competence could be lower because of poor communication and teamwork.

4.5 Data collection

Instruments to collect data were generated and created to provide answers to the following research question: *How does an educational escape room on research integrity affect the motivation of master students towards research integrity?*

The tools were initially designed to provide an answer to the following sub-questions:

How are master students motivated during the escape room activity on RCR?
How do escape room elements affect the intrinsic motivation of master students?

The answers to the sub-questions will generate insights in answering the research question.

The tools used and created for this study are: an *observation scheme* on behavior that indicates (intrinsic) motivation, a *focus group instrument* focused on motivational aspects and empowerment cognitions (see section 3.3) and a *post activity survey* all focused on the same motivational aspects. These motivational aspects are *Interest/Enjoyment*, *Perceived competence*, *Effort/Importance*, *Pressure/Tension*, *Perceived choice*, *Relatedness* and *Value/Usefulness*. These motivational aspects were used from The Intrinsic Motivation inventory (IMI) (Center for Self-Determination Theory, 2021; Ryan, 1982). Gathering more information regarding these motivational aspects will result in more information on intrinsic master student motivation during an educational escape room on RCR.

According to IMI scale description, interest/enjoyment, perceived competence, perceived choice and relatedness are theorized to be indicators and positive predictors of intrinsic motivation (Ryan & Deci 1985; 2010; 2017; Center for Self-Determination Theory, 2021). Pressure/Tension then is a measure that is theorized to be a negative predictor for intrinsic motivation. Effort/Importance is mostly a separate variable for motivation, with potential viable input for intrinsic factors and value/usefulness is used for determining (goal) internalization on self-regulation. Information about these specific categories could be useful for predicting intrinsic motivational aspects, but the participants might be intrinsically motivated for other reasons and factors than the escape room is aiming for. For example, when a participant is putting a lot of effort in completing the escape room because he or she enjoyed the activity on RCR and research integrity discussions, then his or her motives are more intrinsic towards research integrity. But when a participant put a lot of effort into the escape room just to complete it in time and to solve puzzles, then someone is not intrinsically motivated towards research integrity. Then someone is intrinsically motivated because he or she likes to solve puzzles, regardless of the topic.

So, the three instruments were designed to provide answers to the research question and sub-questions. For example, observations during the activity and the focus group discussions of the first group indicated that two out of four participants of the first group felt distant to their teammates (which suggests negative relatedness). Then when analyzing the post activity survey results, the motivational category *relatedness* scores relatively low for two out of four participants. This survey score confirms the relatedness indications collected in the observations and the focus group. Since all three instruments tried to point towards the same result, the outcomes should be reliable to use for answering the research question and sub-questions. Therefore, these three methods of collecting data tried to increase the validity and reliability of the research because of data triangulation (Carter, Bryant-Lukosius, DiCenso, Blythe, Neville, 2014; Polit & Beck, 2012).

The first instrument is an observation scheme, consisting of motivational codes (Appendix A). Appendix B gives an example of how these observations were noted. During the pilot and the test rounds, the participants were observed and moderated together with master student Sarah Seghier. The observations served as possible confirmations for the statements made in the focus group and the survey after the escape room activity. Moderating with two moderators also increases the reliability and validity of the observations. Secondly, a post activity survey was created (Appendix C). This survey consists of altered questions from the Intrinsic Motivation Inventory, which is originated by Ryan (1982) and is constantly being altered by many researchers. "The Intrinsic Motivation Inventory (IMI) is a multidimensional measurement device intended to assess participants' subjective experience related to a target activity in laboratory experiments" (Center for Self-Determination

Theory, 2021, p. 1). The instrument addresses seven scales/factors that are used in the measurement of intrinsic motivation (discussed earlier in this section). For every scale, 2-3 of the questions were selected (based on usefulness for the study) and slightly altered towards the research integrity theme. For example, *I enjoyed doing the activity very much* became *I enjoyed doing the escape room activity very much*. Slightly altering the questions towards an escape room activity tried to maintain the validity the existing questions had. A question selection was needed, because else the survey would have been too long for the participants. A survey which takes too long to fill in will only lead to irritation among the participants. Most categories consisted of six or seven possible questions in the inventory to use in a survey on intrinsic motivation. Almost all questions had a positive and a negative question possibility. For example, *this was a fun activity to do* (positive) versus *I thought this was a boring activity* (negative). When two questions aimed at the same subject, one of the two was selected for the survey (which was mostly the positive question). Thirdly, a focus group instrument was created (Appendix D). This instrument consists of questions about decisions made during the escape room activity. The created questions were a mix of questions related to motivation and engagement, and questions related to empowerment and awareness towards research integrity and the escape room activity. Lastly, half of the escape room activity was piloted on research master students. Observing the pilot provided insights in the functionality of the escape room activity in Microsoft Teams and PowerPoint. Also, feedback and input from the research master students was very important in redesigning the escape room activity.

5 Design process

5.1 Design pre-pilot

The escape room design features multiple cases and scenarios based on real-life academic contexts. Most of the situations, dilemmas and scenarios are based on experiences that students have in their own research contexts. The emphasis during the escape room activity was on the grey dilemmas in research practices. The free-rider dilemma, mentioned in the introduction, was used as a grey dilemma in the escape room activity. Another example of a grey dilemma that was used in the escape room activity is: what to do with flawed data and not enough time left to redo your research? The goal of this escape room activity is to make master students more aware of grey dilemmas and unforeseen scenarios in conducting research. The escape room activity could then empower and motivate the students to gain more confidence and insights on how to deal with certain grey situations in their own research.

The escape room consists of four rooms and four dialogues in between the rooms (see figure 5 for a visual overview of the escape room structure). The dialogues, dilemmas and rooms are further discussed in section 5.1.1 (rooms) and 5.1.3 (dialogues and dilemmas). The participants were able to click on all sorts of objects inside the room to solve puzzles. The arrows in the dialogues granted the participants the option to navigate freely throughout the dialogue. For example, they could replay the dialogue if it was not clear enough the first time. The puzzle structures used in all rooms are open (see figure 4). There are 2-3 puzzles per room which all provide a

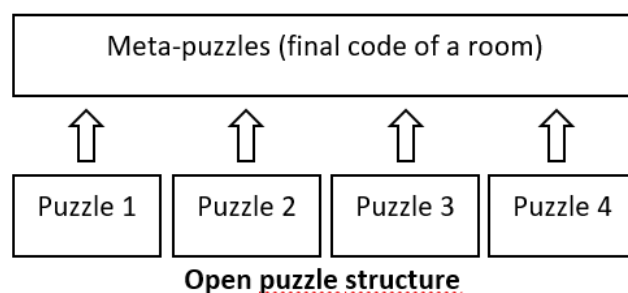


Figure 4. Open route puzzle model, adapted from Nicholson (2015)

piece of the final code. The puzzles can be completed in any order to eventually escape the room. Even though open puzzle structures are used the least in escape rooms globally (Nicholson, 2015), the open structure has great potential for motivating and empowering master students. From all puzzle structures, the open structure provides the most potential for student autonomy (freedom of choice) inside the room. For empowerment, an open puzzle structure leads to a feeling of more (perceived) control. The students are able to solve every puzzle at the same time and they are able to choose between two puzzle routes. They are almost in full control of their choices, how they decide to play and when they decide to progress (excluding the time limit).

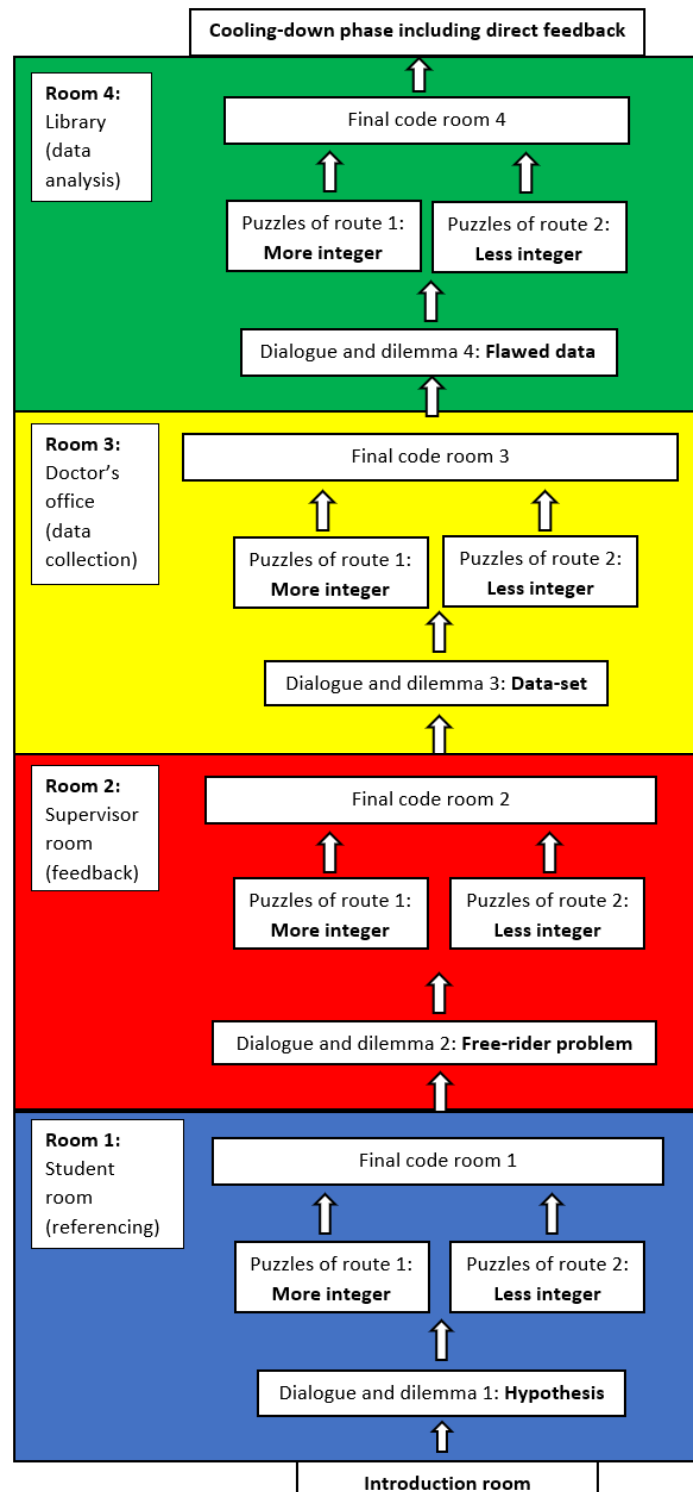


Figure 5. All escape room parts in correct order with a bottom-up structure. The ER activity starts with an introduction room, which is played as a separate activity before the official escape room activity to show participants the controls and how the ER works in PowerPoint. Every room of the ER consist of one dialogue and one escape room part on grey dilemmas in research integrity and on how to conduct RCR. Inside the escape room part of a room (routes after the dialogue), participants are able to follow two puzzle routes inside the room towards the final code of the room. Both routes will give the code, but one of the routes is more integer than the other. The twists to the dilemmas are integrated in the 'Dialogue and dilemma' blocks.

5.1.1 Rooms

The escape room activity consists of four rooms: a student room, a supervisor room, a doctor's office and a library. The rooms represent all a different phase from the writing process of a thesis or project. To introduce the theme and the narrative to the participants, an introduction room was designed as well. Participants played this room before participating to the actual escape room activity.

The *introduction room* consisted of a simple escape room without research integrity contents (see figure 6). The room was created for participants who have never (or not that many times) participated in escape rooms before. The room also provides explanation on the controls inside the escape room in Microsoft PowerPoint. Furthermore, background information on research integrity and RCR is provided regarding the narrative and goals of the activity in other slides of the introduction PowerPoint. An introduction was implemented because it will help the students to progress more fluently through the escape room and to make clear to them why they are doing the activity in the first place. An introduction with clear goals will also aid in creating meaningful play (see section 3.4).

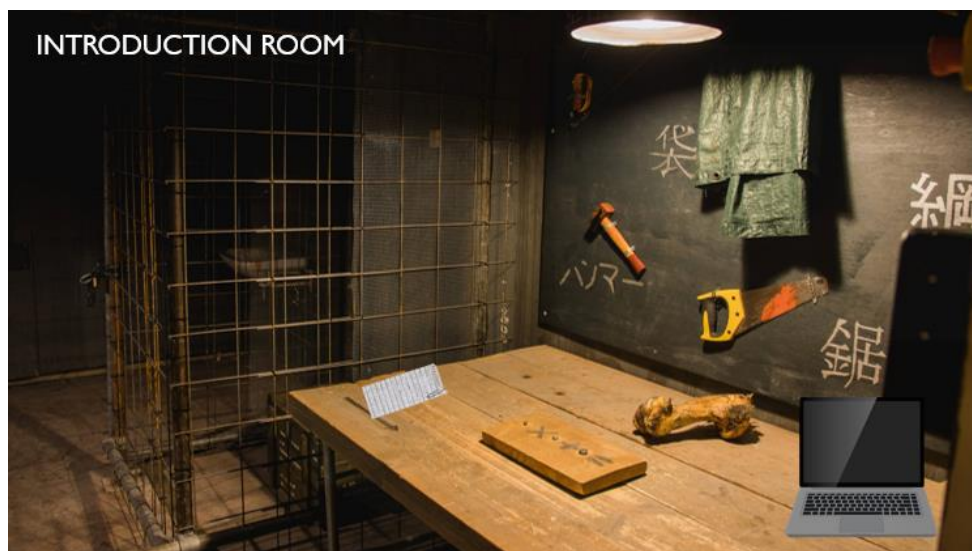


Figure 6. The introduction room

The *student room* represents the referencing phase of writing the thesis and is the first room of the escape room activity (see figure 7). Referencing is a skill all students and researchers should be able to carry out responsibly. The in-game introduction, theoretical background and method section of the thesis have already been written. Now the fictional students need to reference properly to the information they used for creating these sections, before progressing to the second room. Inside the room the participants needed to use their prior knowledge on how to reference to other studies responsibly (or what they believe is responsible). The participants are able to choose between two puzzle routes, with one puzzle route more integer than the other. The participants then need to choose a route which will indicate, along with other factors, their understanding on research integrity and their awareness on complex integrity dilemmas. Also, time pressure is always a limiting factor in conducting research, so every room has a time limit based on difficulty to make the context even more recognizable.



Figure 7. The student room (referencing)

The *supervisor room* represents a first feedback phase on the fictional students their process and is the second room of the escape room activity (see figure 8). The fictional students get feedback from their supervisor to go over before they can conduct their research. Inside the room participants once again need to use their prior knowledge about conducting research, but this time it is focused on the use of feedback from external parties in own productions like a thesis. The fictional students have to make a decision on how to process the feedback they got from the supervisor in their own thesis to eventually progress to the third room.



Figure 8. The supervisor room (feedback)

The *doctor's office* represents phase of data collection and is the third room of the escape room activity (see figure 9). The fictional students got the thumbs up from their supervisor to conduct their research. In the room the participants are collecting and completing the research data by solving puzzles and working together as a team. They are able to choose between an easy, but less integer route by choosing an already existing data-set which is slightly flawed or a more integer but also a more difficult and time-consuming route by completing their own data.



Figure 9. The doctor's office (data collection)

The *library* represents the phase of data analysis (see figure 10). The fictional students got their data from the doctor's office and they now need to perform a statistical analysis to process the data and complete their thesis. The participants are once again able to choose between two routes, one easier but less integer route and one harder but more integer route. The time limit played an important role in the decision-making regarding all the rooms.

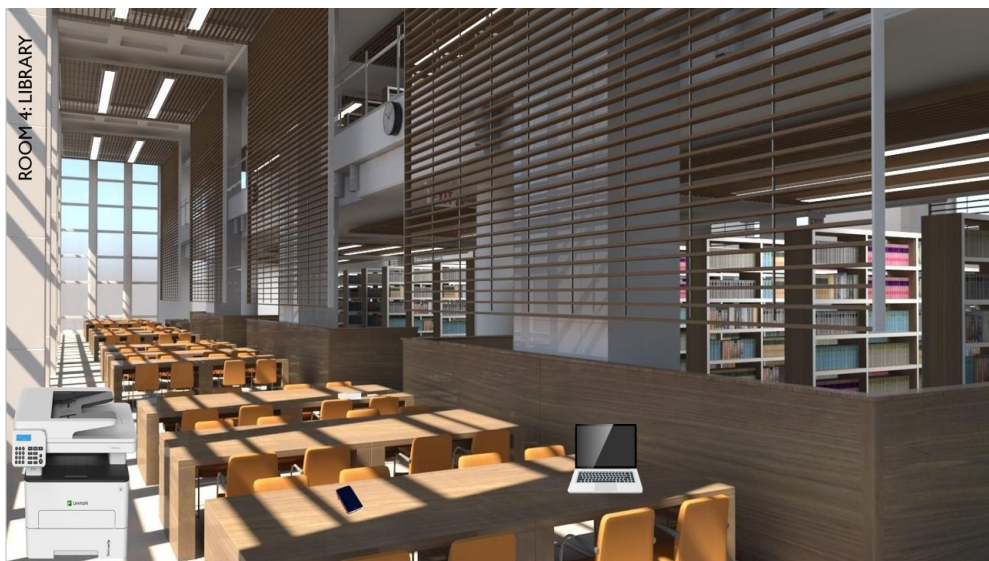


Figure 10. The library (data analysis)

5.1.2 Routes

Inside the rooms in PowerPoint, participants were able to click on all kind of objects or areas. Every room contains a laptop or computer. The computer held the information that was needed to progress through that particular room. This information was hidden in one of the application symbols on the desktop, which contains a form with questions about the room. The participants provided answers to the questions by solving puzzles and deciding what to do when encountering grey dilemmas. The answers they gave were linked to a number, which together provided the code to complete the room.

The participants had to choose between two routes to complete the rooms. One of the routes is more integer than the other, but takes more time to complete. The other route takes less

time, but is less integer. The goals of choosing between two routes is to see if the participants are aware of the different routes while participating, and if they are aware of the integrity difference between the routes. An example of the two available routes in the doctor's office: the participants could decide to complete the first route, which is conducting their own data. In the third room, the hardest room, it is tough to complete the own data within the time limit of 15 minutes. It is possible with good communication and task division, but even then it remains rather difficult. The participants also could choose the second route: the desk contains the data files mentioned in the dialogue. These files consist of 20 participants with complete data. The code that is noted on the files is also the escape code for the third room. This option however is less integer than the other option, because the data is slightly flawed. Time pressure could cause the participants to choose route 2 over route 1 or they do not even notice the different routes at all. If the participants want to stay as integer as possible they need to choose the most difficult route, even though this might take more time than they have left.

5.1.3 Dialogues and grey dilemmas

In between the rooms, one dialogue per room is created to strengthen and visualize the narrative for the participants. The dialogues bring the main character James, Justin and Amber in different research situations, which are connected to the escape room that follows after. These dialogues are concluded with a grey dilemma, where the participants need to choose one answer out of four options. After the decision was made, a pressured twist of the dialogue appeared. For example, the supervisor is (unintentionally) pressuring the fictional characters to act less integer because of a closing deadline. The goal of the twist is to make the participants aware of the pressure and of the fact that changing the answer because of pressure, might not be the most integer thing to do. The decisions that the participants make in both the dialogues and in the rooms do not affect the narrative.

The first dialogue and dilemma is about the *hypothesis* (see figure 11). Does the group wants to keep the hypothesis they constructed at beforehand, or do they want to wait on the research results? This dilemma was based on the questionable research practice *HARKing* (*hypothesizing after the results are known*). In short, HARKing is defined as "presenting a post hoc hypothesis in the introduction of a research report as if it were an a priori hypothesis" (Kerr, 1998, p. 197). HARKing is a QRP because it is not considered as misconduct, but the behavior is frowned upon by other researchers. HARKing in general does not damage the integrity of the research process directly, but lacks transparency (Murphy & Aguinis, 2019). HARKing is a grey area in research, which makes an issue on HARKing a suitable grey dilemma to use in the escape room on research integrity.

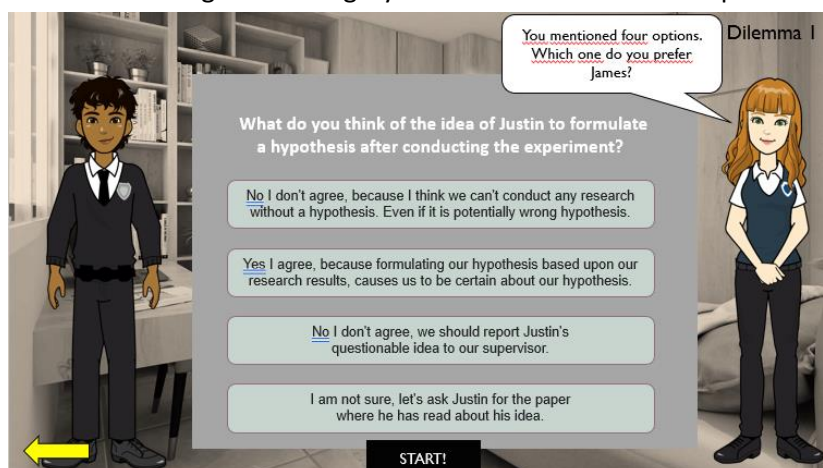


Figure 11. Hypothesis dilemma

The second dialogue and dilemma is about *teamwork* (see figure 12). What to do when a group member is free-riding? As explained earlier, the free-rider problem is defined as a situation where a group member contributes “little or nothing towards the cost of the good, while enjoying its benefits as fully as any other member of the group” (Kim & Walker, 1984). The free-rider problem is a problem in group dynamics in general. The problem might occur in every stage of education from middle school to higher education. However, not every student really knows what to do when encountering the free-rider problem. The free-rider problem makes a fine example of a *grey dilemma* in research, because there are several viable options that are able to deal with the problem.

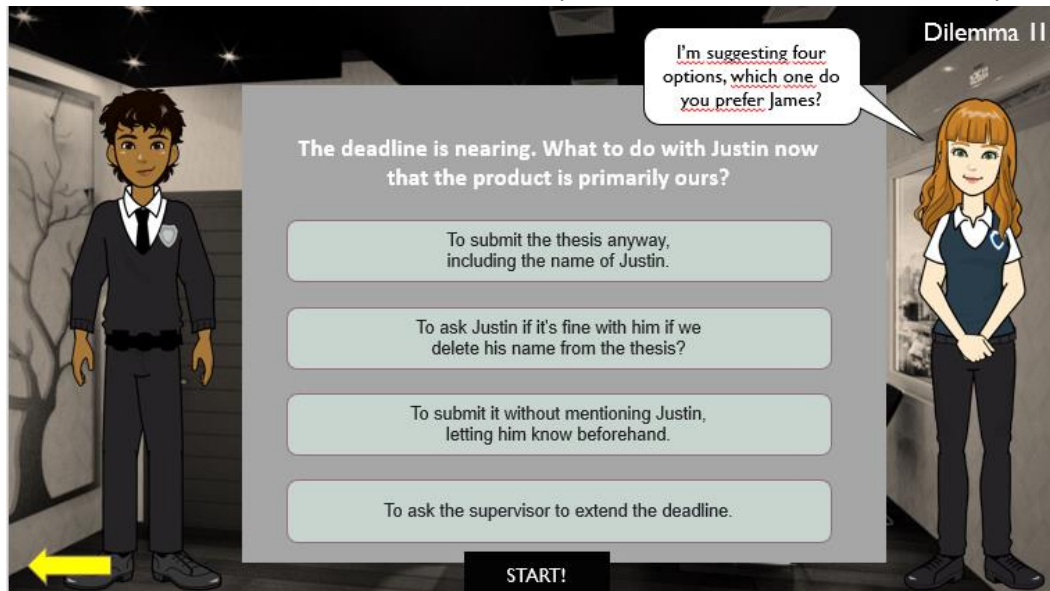


Figure 12. Teamwork dilemma (free-rider problem)

The third dialogue and dilemma is about a given *data-set* (see figure 13). The fictional students are just about to conduct their research. Then they hear from their doctor supervisor that there are only twelve patients coming to conduct the BMI research on (which makes the sample size too little to conduct quantitative research). The doctor then offers a solution, a data-set of 20 patients created by a colleagues intern student. The only problem is that the methods and the logbook of the data-set was missing. Choosing the given data-set is the most easy option in this dilemma, but it is also possibly a questionable research practice. It is of course a possibility to retrieve the methods and logbook from the colleague, but even if they retrieve these important parts the behavior could still end up being questionable. The data-set might have already been used by the student that conducted the research in the first place, which would have led to the use of one data-set in two publications. This type of research behavior is an example of a QRP (Steneck, 2006). So the students need to be aware of two different aspects in this dilemma to make a well-thought decision.



Figure 13. Data-set dilemma

The fourth dialogue and dilemma is about *flawed data* (see figure 14). What to do when the only data that is collected for the research is flawed, with only little time left until the deadline? The use of flawed data in a research production is at first something that occurs almost in every type of research. It is not necessarily bad to use flawed data in a publication, as long as the publication is transparent about the flaws in the data or research methods. When a publication uses flawed or bad data without being transparent about these flaws, this is defined as misconduct (Anderson et al., 2013). There is still debate about if flawed data is a bad thing, which makes this issue a suitable grey dilemma for the escape room.

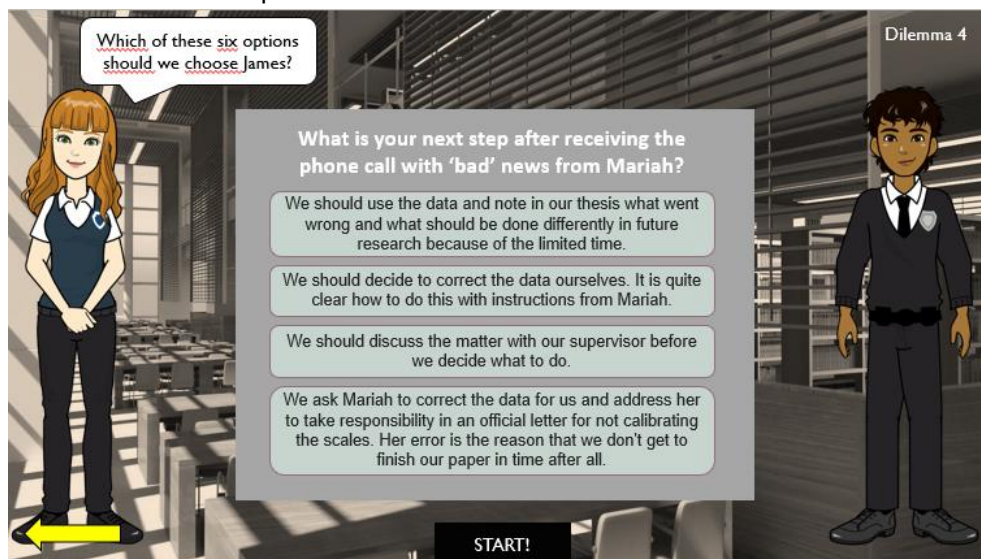


Figure 14. Flawed data dilemma

The participants need to work together and discuss the dilemmas as a group to complete the dialogues and escape rooms. When the four dialogues and escape rooms are solved, the escape room activity is completed.

5.1.4 Storyline/Narrative

The in-game narrative consists of three master students that are working on a thesis together. In the escape room activity, the three master students are James, Justin and Amber. The participants of the escape room activity will assist the main character James in making decisions, completing tasks and solving problems and puzzles in the thesis completion process. The theme of the escape room is RCR, so the focus is on conducting research responsibly rather than experiencing spectacular cases of misconduct (more positive than negative). The subject of the thesis is the Body Mass Index (BMI). It is a subject that everyone knows and is therefore recognizable for students, but other subjects might prove just as suitable. The context of the escape room was designed to be as recognizable as possible for students regarding their own academic research practices. A thesis is an assignment every student has to complete at the end of a bachelor study. So every master student has experience with writing a thesis, which potentially makes the activity meaningful and valuable to them. When an activity is meaningful and valuable to someone, the motivation to continue with the activity is most likely increased (Frey & Fisher, 2010; Ling, 2018).

Before the activity started, the participants got some background information about the fictional master students and their thesis. This background information was about the narrative, but also on how an online escape room in PowerPoint works. This introduction was created in a separate PowerPoint-file. The introduction prepared the participants to what they had to do in order to complete the activity as a whole. The research integrity theme was explained, but of course without telling the students that they had to be aware of grey situations and dilemmas (because that factor was observed during the activity and discussed in the cooling-down period). The participants received the introduction escape room at forehand as preparation for the activity. The introduction explained that the fictional students had already been actively working on the thesis. The subject they chose to make a thesis about is: the Body Mass Index (BMI). The three fictive students worked out the introduction, theoretical background and method section of the thesis already. At the start of the escape room activity, the trio just starts to revise the work they have done already. Before making arrangements with their supervisor to conduct their research (room 2), the hypothesis needs to be revised (dialogue and dilemma 1) and the referencing of the introduction, theoretical background and method section needs to be done (room 1). The introduction and instruction intended to contribute to the feeling of meaningful play by clarifying the goals of the escape room (Salen, Tekinbaş & Zimmerman, 2004; van Bree, 2011; 2013). A more practical explanation on the narrative/storyline is stated in appendix E.

Summarizing, the dialogues, dilemmas and rooms are designed to be connected to each other. The rooms represent different phases of research, whereas the dialogues and dilemmas are contextual situations that might occur while conducting that certain phase of research. The first dialogue and room represent a *preparation* phase, by gathering information, revising the hypothesis and referencing to the introduction, methods and theoretical framework. The second dialogue shows the free-rider problem, one of the students did not contribute as much as the other students on creating the introduction, theoretical framework and method section. The second room represents a *feedback (revision)* phase on the created sections, with feedback from the supervisor. The third dialogue and room represent the *data collection* phase of conducting research. The dialogue is an introduction on the room with a data-set dilemma. Then to complete the third room, the data-set needs to be completed by solving puzzles and tasks. The fourth and last dialogue and room represent the *data analysis* phase. The fourth dialogue is connected with the third room, because the dilemma is about the data-set being flawed. The room then focusses on analyzing the (corrected) data-set in Statistical Package for the Social Sciences (SPSS). To complete the room, students need to solve puzzles in order to obtain the numbers needed to complete a SUM-score in SPSS. The escape room

as a whole is completed when the final code of room 4 (SUM-score) is used to print the thesis on BMI. Creating a connection between dialogue, dilemma and room will cause more synergy between the different aspects. When the dialogue and dilemma are too distant from the contents of a room it should be connected to, answering the dilemma responsibly might feel irrelevant or not important for the students. The dialogues, dilemmas and rooms need cohesion to create a fluent escape room experience for the master students.

6. Pilot

The pilot was useful to test the design on functionality and to generate redesign insights. The pilot group consisted of three master students and one master teacher, with another master teacher spectating. The pilot tested two out of four designed rooms: the doctor's office (third room) and the library (fourth room). These rooms were chosen because the third room was the most challenging one to solve, and the last room should be easier for the participants to create low chance of the participants failing the last room. Failing the last room could cause the participants to feel like they have failed at the escape room activity as a whole, which influences their perceived competence negatively.

During the pilot, the activity was mediated and observed together with master student Sarah Seghier. All the participants together with the moderators were present in one Microsoft Teams meeting. The participants put on their cameras as they downloaded the PowerPoint-file before opening it in the desktop-application. Every participant got their own PowerPoint-file to download. One of the participants got the control of the decision making (version 1 of the PowerPoint-file), when discussing the dilemmas and filling in the laptop forms inside the rooms. All participants can progress by clicking an answer in the dilemmas or filling in the code of the room (version 2 of the PowerPoint-file), but the decisions the leader made were observed by the moderators to use in the cooling-down period. Data from the pilot was collected, but not used and published in this study. The data was not representative to the data collection of the test rounds, because only half of the escape room activity was piloted and the design changed significantly after the pilot evaluation.

6.1 Pilot evaluation

The participants of the pilot enjoyed the escape room activity and were impressed by the design. They recognized the dilemmas and pressure during the activity, because they could relate them to their own research experience. The participants mostly thought they made the right choices for themselves, which characterizes confidence and knowledge about research integrity and how to make decisions when encountering difficult dilemmas while conducting research. They also learned more about the type of dilemmas and that not every issue is easy to deal with on their own. Not every issue is just answered with one right or one wrong answer, which granted the participants more insights on grey dilemmas in research practices.

The online collaboration and communication was hard sometimes, the participants could not see each other (unless they had two screens) and did not know exactly what everyone was doing (because the escape room was played in full-screen mode). While encountering the dilemmas inside the dialogues, the participants were discussing and thinking about research integrity. When entering the rooms however, their focus altered. The participants were focused on completing the room rather than choosing the most integer puzzle route. So, the pilot execution provided interesting results. These results granted more insights in what to adjust for the final design (see table 2). Also feedback and comments from the teachers and students helped a great deal.

6.2 Escape room redesign

In the redesigning phase, the activity was adjusted in several ways to counter technical issues as much as possible and to transfer the research integrity experience as smooth, complete and clear as possible. An example of the changes that were made, is that all the rooms and the dialogues got separate timers. The timers were calculated on the size (amount of steps that have to be taken) and difficulty of the rooms. For the dialogues they are in all four occasions the same (see table 2).

The data collection started after the pilot. The escape room activity on research integrity was redesigned based on the insights, granted by the pilot experience (see table 1). The activity then was tested on three groups of four master students from different master studies.

Elements in need of improvement	Improved and added elements
1. The discussions during the dilemmas were too long. One of the discussions took almost 15 minutes. The discussions are part of the experience of course, but apparently some dilemmas were very interesting (which is a good thing). If all four dilemmas took 15 minutes of discussion time, the whole experience would last for two full hours (excluding the cooling-down period). This is extremely long for an educational activity.	Added timers to the dialogues and the rooms. Clicking the START! button starts the timer (see table 2).
2. The direction inside the rooms was too vague. The participants did not know at first what the room was expected from them to do in order to complete it. For example, it was not entirely clear that the participants had to choose between two viable routes inside the rooms. This was intentional at first (because the master students needed to be aware of the two routes that differ on integrity), but the distinction between the routes needed to be more clear in order for the master students to recognize the intention.	Google Forms was replaced with macros in the Windows version and the forms got more guidelines for the participants. Every puzzle got a specific name and number, that matched with the routes in the form on the laptop (which is present in every room).
3. The difficulty of the doctor's office (third room) was too high and because the direction was not clear enough, the participants did not even come close to the final code.	The third room got more guidelines and the puzzles were made a little easier to solve.
4. A MacBook version for PowerPoint is needed because a MacBook device is compatible with most add-ins (like macros) in PowerPoint.	An extra version for MacBook users was made because of incompatibility with macros. The MacBook version was made non-interactive (regarding macros) to smoothen the experience.

Table 1. A summary of the pre-pilot design elements that were in need of improvement (left column) and the actual improvements and changes made during the redesigning phase (right column).

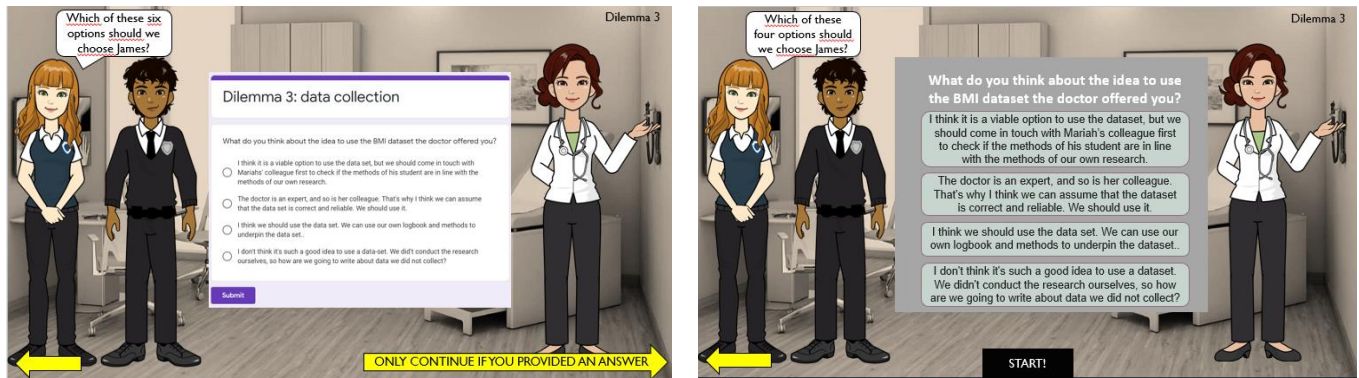


Figure 15. An example of the ER redesign. The figure shows the dialogue design of the doctor's office (third room) pre-pilot (left) and post-pilot (right). The google forms got replaced with clickable macro options and a timer was added to limit the discussion size to smoothen the escape room experience.

Escape room element and difficulty	Timer
Dialogue 1: Hypothesis	5 minutes total 3 minutes for the first dilemma 2 minutes for the twist
Room 1: Referencing (easy)	10 minutes
Dialogue 2: Teamwork	5 minutes total 3 minutes for the first dilemma 2 minutes for the twist
Room 2: Feedback (medium)	15 minutes
Dialogue 3: Data-set	5 minutes total 3 minutes for the first dilemma 2 minutes for the twist
Room 3: Data collection (hard)	15 minutes
Dialogue 4: Flawed data	5 minutes total 3 minutes for the first dilemma 2 minutes for the twist
Room 4: Data analysis (easy)	10 minutes

Table 2. Escape room elements with room difficulties and their timers.

7 Results

The data-collection was aimed towards providing an answer to the following sub-questions:

How are master students motivated during the escape room activity on RCR?

How do escape room elements affect the intrinsic motivation of master students?

Answers to the sub-questions will provide useful insights and information to eventually answering the research question of this study:

How does an educational escape room on research integrity affect the motivation of master students towards research integrity?

Three groups of four participants each participated to the RCR escape room. The data from three test rounds was collected by the use of an observation form (appendix A and B) focus group instrument (appendix D) and a post activity survey (appendix C). The used methods are aimed to ensure data-triangulation, by focusing on answering the same research questions. The three test rounds took place on respectively the 23rd of December (first group), the 5th of January (second group) and the 19th of January (third group). The performance of the participants was recorded within Microsoft Teams, which granted the opportunity to observe the activity afterwards. During the escape room, the two moderators provided hints for the participants when the whole group got stuck in a room. The three data collecting methods are all based on the same motivational elements (explained in section 4.5): interest/enjoyment, perceived competence, effort, value/usefulness, pressure and tension, and perceived choice (autonomy). The interpretation of the results regarding these motivational elements are showed for all groups per element. The survey results per group are presented in charts and participants' quotes from escape room participation and the focus group are used to clarify interpretations on the motivational elements.

7.2 Interest/Enjoyment

The interest/enjoyment category, is the first of seven elements that measures intrinsic motivation towards the escape room on RCR.

Observations of the three groups, focused on signs of interest were mostly showed when the groups discussed the grey dilemmas in the dialogues. The participants of all groups discussed the first two dialogues and grey dilemmas intensively, but the participants felt like the dialogues and the rooms were separate activities from each other. They thought that there was just too little cohesion between the contents of the dialogues and the escape rooms. This caused interest towards research integrity while discussing the first two dialogues and grey dilemmas (HARKing and free-rider problem), but interest declined towards research integrity while completing the four escape rooms. The size of the dilemma discussions decreased towards the third and fourth dialogue. The interest towards the integrity dialogues and dilemmas faded after they got the feeling that the type of answer they chose, did not influence the escape room progress at all. A few participants even spoke out loud, saying that they did not care about the answers that they would give on solving dilemma three (data-set without methods). The third group showed the most interest while discussing the grey dilemmas. However, inside the rooms they were mostly occupied with completing and solving puzzles rather than discussing which puzzle route was the most integer. All groups did not recognize the function of the puzzle routes per room at first. They thought that both puzzle routes needed to be completed in order to achieve the final code to escape the room. Every room consisted of two puzzle routes, with one *less integer* puzzle route and one *more integer* puzzle route. The first group did not even know there were two different routes in the first place. The second and third group figured the functionality of the two routes out after progressing through the first two dialogues and rooms.

The three groups did not show much observable enjoyment, two out of three groups even showed more signs of frustration. Frustration influences the enjoyment of the participants negatively. Two participants from the first group faced several technical issues, which caused their experience to not progress fluently at all. The two participants eventually even needed to spectate the leading participant to participate in the first place, because the ER PowerPoint-file crashed or did not start. The main reason for the inconvenience was because MacBook devices are not compatible with the used add-ins in the ER PowerPoint-file (which still remained a problem after the redesign). Next to the technical inconveniences, the participants of the first group were visibly frustrated because they often got stuck inside the first three rooms. They were not able to figure out what to do inside most rooms and the puzzle difficulty seemed too high inside the second and third room. The fourth room went most fluent, because after three rooms of experience they were able to more or less figure out what the room asked them to do in order to complete it. The first group did only complete the fourth room in time, which also could have caused the signs of frustration inside the first three rooms. Two participants of the second group also encountered technical issues, but less severe than the participants from the first group. The second group also got stuck inside the rooms less than the first group. The group was visibly searching for the direction within the first room. After the first room, the direction within the room became more clear to the second group. The second group solved one out of four rooms in time. The third group did not encounter technical issues at all and only got a little stuck inside the hardest room, the doctor's office. All groups got stuck inside the doctor's office and needed hints from the moderators to be able to progress to the next room. The third group completed three out of four rooms in time. Besides the observable frustrations in the first two groups, all groups did try their best to communicate and work together online to eventually complete the escape rooms and to provide well-thought answers to most of the grey dilemmas discussed in the four dialogues.

The focus group results did mostly confirm the observations for all groups. The first group confirmed that they struggled with the direction within the rooms and the puzzle difficulty being too high:

Group 1, student 1: One thing that we struggled with is that the puzzles were then so hard that you kind of just could not get out of them. So we just took the easy route, despite knowing it was wrong. But it was like kind of frustration.

Group 1, student 2: I guess I kind of liked it sometimes, but confusing what to do. But I guess that is always the case in an escape room. Although it was not clear from the beginning that it was kind of wrong or like wrong to take that shortcut. In an escape room you always choose the easiest way because that's how you solve the problem.

Group 1, student 2: I think it went quite well, like it was a bit difficult in the rooms due to people not having their own work and PowerPoint. So we could not see as many things as we wanted at the same time. But, like discussing the dilemmas and stuff went well. But it was like kind of a lot of frustration. Like, what do we have to do to solve the puzzle before they can take a route?

Group 1, student 3: And for me I... maybe it is because it did not work on my computer, but for me it was confusing what was the easy route and what was the more difficult one.

The fourth student then disagreed with the other three participants, which indicates competence towards research integrity:

Group 1, student 4: I mean, the escape room is on research integrity. So I think it is kind of like given because you have these options and usually the easy choice that is mentioned it is like, oh, this is probably not the best way to.

This is an interesting quote, because it suggests that one participant did understand and recognized the difference in integrity between the two puzzle routes compared to the other three group members. Even though the participants of group 1 showed more signs of frustrations than of enjoyment, more positive answers aroused when asking about the activity on itself. These quotes indicate that the participants did in fact enjoy the activity:

Group 1, student 1: Yeah, I think it was fun.

Group 1, student 2: This is fun and a playful way.

Group 1, student 4: I think it was definitely interesting in the discussion part.

The technical issues group 1 endured were limiting their experience and enjoyment too much to use the same file for the other two test groups. The escape room contents did not change, only some add-ins were removed in the MacBook version to smoothen the downloading of the file and clicking in the slides. Therefore, the second group was more positive about the design and the activity in general, which indicates enjoyment:

Group 2, student 1: It was a good experience. It was nicely put together and the clues were very subtle.

Group 2, student 2: Yeah. I think it is a very fun kind of experiment that you constructed.

Group 2, student 3: So yeah, I agree. It was fun and I did not think this would be possible to even do this online.

Group 2, student 4: I agree with the rest. I think it was a nice design/organization, like the PowerPoint was nice to look at.

But also the second group had issues with finding guidelines and the direction in the first three escape rooms, resulting in more enjoyment and enthusiasm towards the final room:

Group 2, student 2: Especially at the last one, I became a little bit more enthusiastic to actually solve the harder question. Because I started to understand what we needed to do in order to solve it. But in the first few escape rooms I actually just was completely in the dark.

The third group was also rather enthusiastic about the activity and did in general not really struggle with what they needed to do in order to complete the rooms:

Group 3, student 1: Very well constructed! Well done guys. When you use this in an educational setting, you got yourself a masterpiece.

Group 3, student 2: When you use this as a teacher, then you are the coolest teacher around! The design was amazing, which is important.

While discussing the escape room during the focus groups in the cooling-down period, the interest group 1 and 3 had was mostly faced towards the design process and not towards the content of the escape room activity (research integrity). The participants were also eager to know more about the aspects of the activity that did not went well, room three in particular. The second group did not show interest in the design or the activity afterwards:

Group 1, student 2: Can we view the results?

Group 3, student 2: How on earth did you guys made this in Microsoft PowerPoint? I really did not know that was possible.

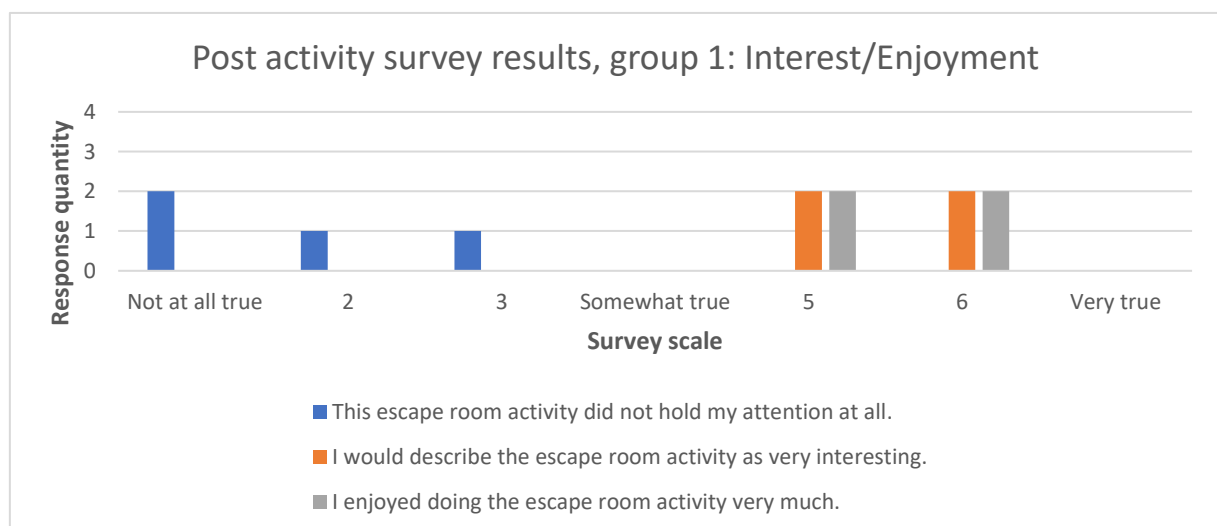
Group 3, student 3: Did you hired some escape room expert for this design?

The third group mentioned in the focus group that they felt like the escape room as a whole was taking too much time to complete (for over one hour). This might have caused participants to lose their focus and concentration which could affect their enjoyment:

Group 3, student 2: I thought the activity was a bit too long if I am honest. Half the activity would have been better.

The survey results of all groups confirmed the observations and focus group information (see figure 16). All groups score rather high (5 or higher on a scale of 7 with one participant scoring 4/7) on a positive enjoyment and interest statement. This confirms that the participants in fact felt like they enjoyed the activity and thought the escape room was interesting. The third statement was stated negatively, asking the participants if the activity was not able to hold their attention at all. According to the three groups, the activity was able to hold their attention (scoring a 1 or 2 out of 7 with one participant scoring a 3 out of 7).

These results on the interest/enjoyment category suggest that all groups enjoyed the activity, despite several signs of frustration. The participants were mostly interested towards RI contents in the dialogues, while discussing the grey dilemmas as a group. However, the connection between the dialogues and the rooms was too little for the participants. This caused the participants mainly to be occupied with completing and solving puzzles, rather than discussing which puzzle route was most integer in their opinion. Also, the direction within the rooms was not always clear to the participants, which also caused frustration and visibly limited the escape room experience for especially the first and second group. This lack of direction could have also been caused by the introduction being too vague or not clear enough. Some participants also thought that the puzzle difficulty was too high in some of the rooms, especially inside the third room (doctor's office). The results overall show that the participants were enjoying completing the rooms a little bit more than discussing the grey dilemmas. However, there were no rewards given when a room was completed (only progressing to the next room). This indicates that the participants of all groups were intrinsically motivated towards completing the rooms.



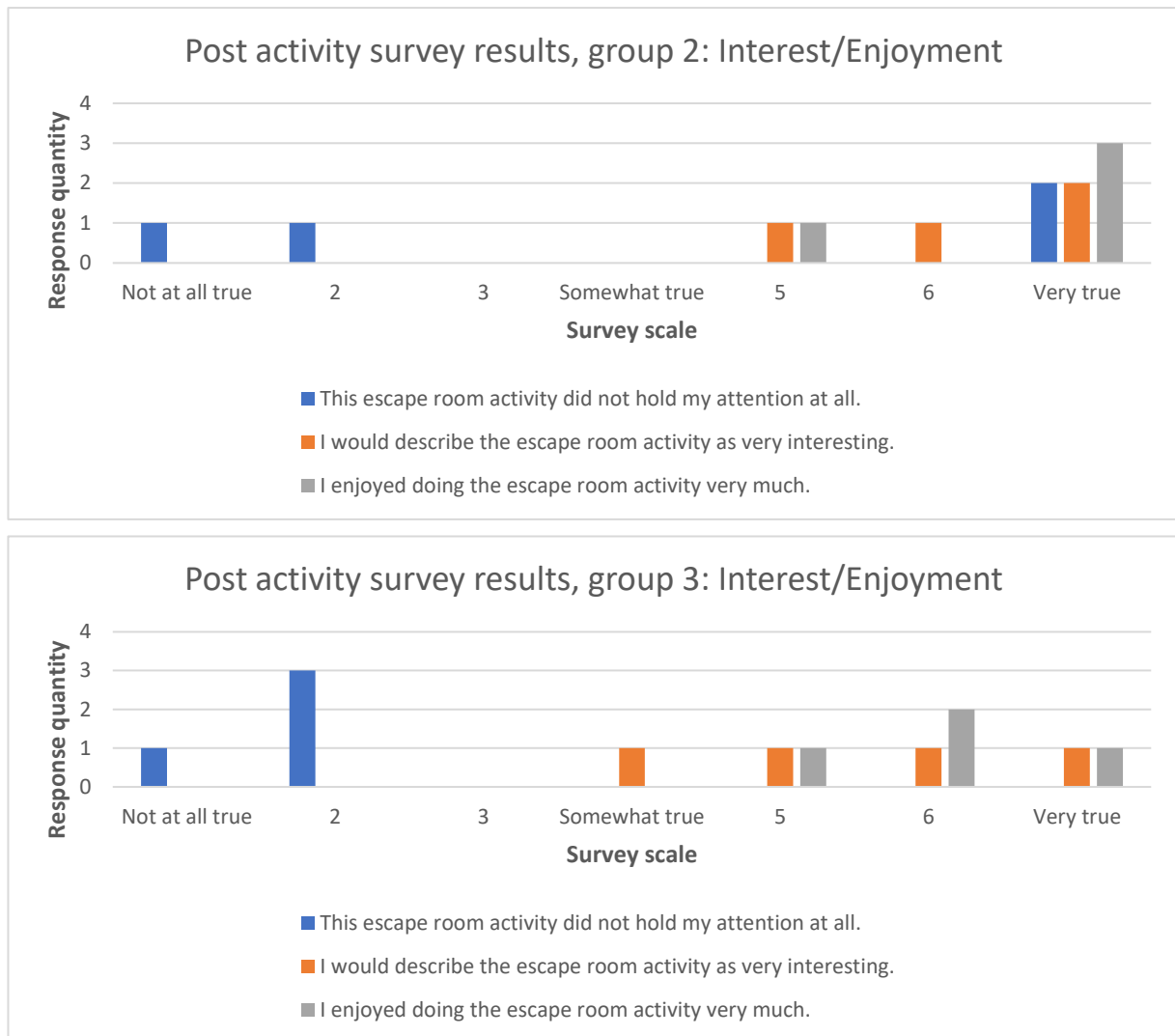


Figure 16. Survey results of all three test groups on the interest/enjoyment category

7.2 Perceived competence

The perceived competence category was measuring if the participants experienced a feeling of competence during and after the research integrity escape room. When the escape room creates a competent feeling, this positively influences the intrinsic motivation of the students.

Observations indicated that the first group showed the least signs of perceived competence of all participating groups. The grey dilemma discussions during the dialogues did in fact show signs of perceived competence, as the participants confidently answered them. In all dialogue dilemmas the participants thought of their right answer option rather quickly and resolute. The decision making was even that quick, that the decision making was almost not discussed and underpinned in the last two dialogue dilemmas (data-set and flawed data). Also, the participants thought of the other answer options as wrong answers mostly. There was of course a difference in integrity between the options, with two more integer options than the other two. However, these were not examples of misconduct. This indicates that the participants did not fully recognize the nature of the grey dilemma. The first group did not understand what they needed to do in the first three rooms. Puzzles were thought to be too difficult and the two routes were too unclear to distinguish while under time pressure. The first group did however solve the fourth room in time, which indicates growth of the group the further they progressed in the escape room. The second group showed more individual

knowledge on research integrity during the escape room. After getting used to the activity in the first dialogue and room 1 (student room), the integrity of the puzzle routes were discussed a little more and the discussions on the dialogue dilemmas were rather elaborate and well underpinned. The puzzle difficulty and the direction of the puzzle routes did however remain unclear for the participants. The third group worked together the most effective regarding all groups. The third group was the only group that knew each other at forehand. The first two groups consisted of participants that were not familiar with the other group members. The discussions on the dialogue dilemmas were elaborate and opinions on answer options were underpinned reasonably. The participants did not always agree with each other, but always found one option that created consensus within the group. The integrity difference of the puzzle routes within the rooms were sometimes recognized, after experiencing the first room. The participants then mostly understood that one route was the easy and less integer option, over the hard and more integer route. However, in all groups there was at least one participant that did not fully understand the two puzzle routes. All groups chose the hard puzzle route to solve in all of the rooms, just as the other two groups did. Choosing the hard and more integer route was mostly caused by participants wanting to complete more puzzles, rather than choosing the more integer puzzle route.

The focus group confirmed most observations on signs of perceived competence, regarding all of the groups. The groups were not fully aware of the puzzle routes functionality and did not think about research integrity as much within the rooms as in the dialogue discussions:

Group 1, student 1: Ooh, so the route numbers were correlated to the easier or difficult one. I did not know that.

Group 2, student 1: And one thing I didn't do is... I didn't really connect to the different routes to the two integrity questions. I didn't realize that most of them also needed an important choice to make.

Group 3, student 3: I think that the integrity difference between the routes needs more clarification at the introduction. I thought that the two routes were both options to take when we could not solve one of them, like some sort of back up.

This lack of a competent feeling towards the research integrity contents, was caused by a few design choices: the missing link between the dialogues and the rooms (which is also discussed in the interest/enjoyment category), working in groups rather than making own decisions, a lack of clarity inside the introduction (stated by the last quote) and the focus group design of the cooling-down period. Firstly, the connection between the dialogues and the rooms was just too little for the participants to know why they are even answering the dialogue dilemmas in the first place. This caused the participants to see the dialogues and rooms as separate activities, which led to more elaborate integrity discussions in the dialogues than within the rooms:

Group 1, student 4: I think it was very divided between the dilemmas and the escape rooms. Because I don't think once you finished the dilemma, you've really connected to the escape room after that.

Secondly, another factor that influenced the perceived competence on decision making, was the fact that the participants needed to work in groups. All groups stated that when they had to make these choices on their own, they would have not always chosen the same option as the group made in several integrity situations:

Group 1, student 4: I think with some of the answers we gave as a group, I might have not completely agreed with them. I might have done differently if I was in the situation by myself. It did not change my opinion on it.

Group 3, student 1: Because it is a game where cooperation is needed, you provide answers your group wants to hear and not particular answers that only you agree on.

Thirdly, the escape room design had not the potential for recoverable loss, which would have given the participants a redo option to decrease the chance of the participants feeling like they failed at the activity. Some participants felt like they failed at the activity, after not completing one or several rooms. The feeling of failure influences the feeling of competence negatively:

Group 3, student 1: We failed the third room and I was like if we failed this room, we are probably also not able to complete the fourth room. I lost courage. Eventually, we completed the room but you just looked a little like a fool. Like we just were not able to do anything right.

The final escape room design aspect that did not fulfill the competence needs of the participants, was the design of the cooling-down period. The participants were eager to know if they did good or if they picked the responsible answers and puzzle routes. The cooling-down period only partially provided answers and confirmation on their questions and uncertainties. This design choice caused the participants to not really know if they did well or not, leading to them being able to recognize grey dilemmas but not learning how to act when encountering them in own research practices. So, this design choice positively influenced their competence in recognizing grey dilemmas, but did not influence their feeling of perceived competence on dealing with them:

Group 1, student 2: Because I was just discussing things that you find interesting, but then the players might not get a final conclusion. It might sound stupid, but it kind of feels that we don't really learn anything about religious integrity. When every choice is 'correct' it is still unclear how we have to deal with the situations presented to us in this specific scenario.

Group 3, student 1: I would be able to recognize certain integrity dilemmas, but I would not be able to act properly. I do not know what the good thing is to do.

Group 3, student 3: I agree, you want to discuss your decisions and answers to the dilemmas. You are looking for confirmation, because there was not much contradiction between the participants while making decisions.

Besides the escape room design choices that influenced the perceived competence negatively or not at all, some participants did feel competent enough to deal with these grey situations in their own research as they were satisfied with their performance during the activity:

Group 1, student 3: I have one thing to add. I thought the activity was a complete overview of integrity. So the experience gave me an overview of what elements integrity contains and what choices I could make when encountering them. That is something I learned.

Furthermore, all groups showed a similarity in learning one specific aspect: that it is always a viable option to discuss dilemmas, issues or complex situations with other students or with your supervisor rather than dealing with the situation alone. This statement made by all groups, shows a sign of increased competence towards reacting on unforeseen situations like grey dilemmas:

Group 1, student 4: I think in a couple of the questions, there was always the option to just ask your supervisor, I guess, which is never the wrong answer.

Group 2, student 3: To me this escape room was a good reminder that if you do not know something, you should just ask people. That is to me personally.

Group 2, student 4: I am a person that tries to solve everything on her own. So maybe I should ask more.

Group 3, student 2: I think that by discussing decisions and dilemmas with other people, you are more capable of making well-thought out choices. Because when encountering a grey dilemma, it is dependent of who you are working with. You could learn new things because of different perspectives.

Group 3, student 4: This experience makes it easier to me to discuss integrity issues with a group or other people. Because you discussed a lot of issues already, why not discuss another one?

Some participants also showed confidence in their grey dilemma decision making. This confidence could have been caused by participants that already made life decisions about (research) integrity matters for themselves, before they participated.

Group 1, student 1: I guess for me it has not really changed because I already had some courses on this.

Group 1, student 2: Yeah, I guess also because you do not have a lot of time to discuss things. So you go with, like what you already think is important. So you do not really come to a new conclusion.

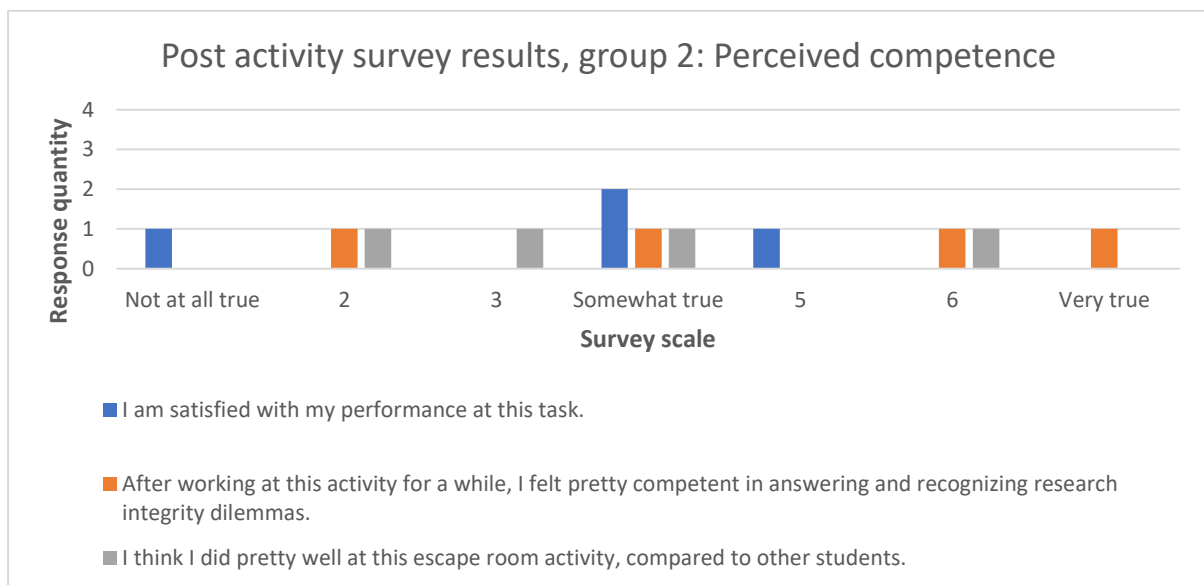
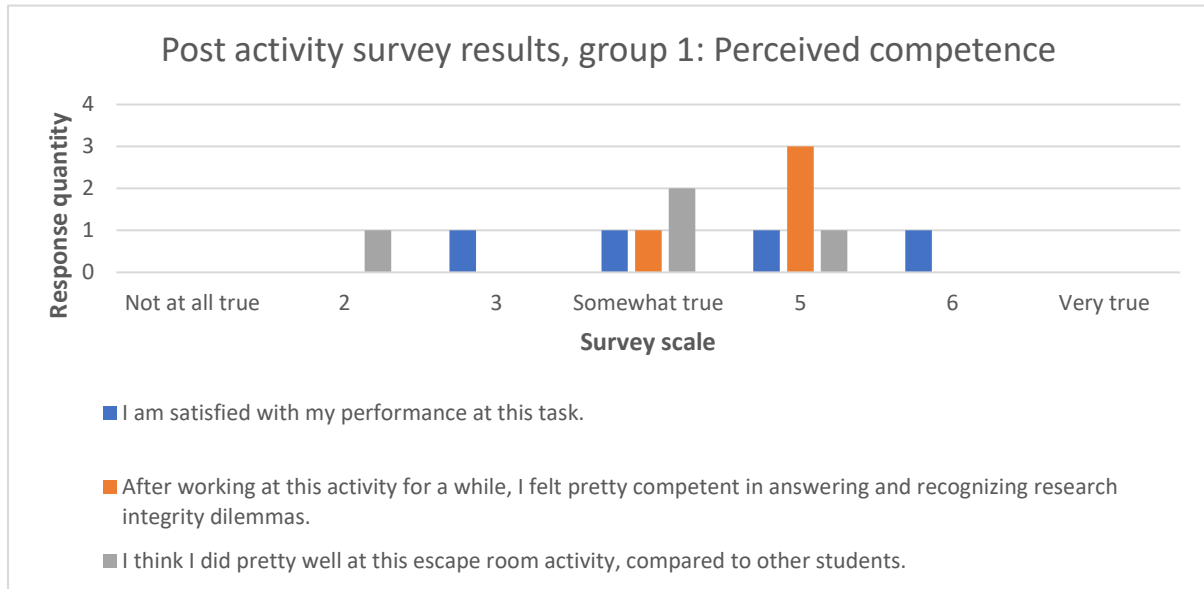
Group 2, student 2: I have come across this kind of integrity questions and problems before. I think that especially time pressure plays a role often because most people know actually what is the right thing to do.

Group 3, student 3: When you did not participate to this activity, would you think different about integrity issues? Because I don't think that is the case for me.

The survey outcomes on the perceived competence category are scattered all over the 1-7 Likert scale (see figure 17). The first and third group scored higher collectively than the second group. 11 out of 12 participants scored a 4 or higher on the scale, which indicates that they felt like they were at least somewhat satisfied with their performance. One participant scored a 1/7, which indicates that the participant was not satisfied at all with his or her performance on the activity. 11 out of 12 participants also thought that they were able to somewhat recognize grey dilemmas and felt pretty competent in answering them. This score might have been higher when the statement was only focused on recognizing the dilemmas, because the observation and focus group results indicate that the participants felt competent in recognizing the grey dilemmas better than being able answering them responsibly. The 12 participants also scored mostly around a 4/7 on the third statement, with two participants of all groups thinking they did not perform better at the escape room than others. This feeling could originate from the fact that a few participants had to participate by spectating others, rather than playing in their own PowerPoint-file. This caused a limited escape room experience, and negatively influenced the perceived competence towards the escape room.

Overall, the results indicate a rather positive influence on perceived competence towards recognizing grey dilemmas and unforeseen situations. The participants learned from the activity that discussing issues or complex situations they encounter in their own research practices with other

students or their supervisor, is a viable option next to dealing with them alone. So, the participants are more competent in reacting and recognizing to unforeseen situations in their own research practices, but they did not feel more capable of solving these unforeseen situations on their own. The escape room experience lacked confirmation for the participants in their grey dilemma decision making, which did not influence their perceived competence (which was not the intended result).



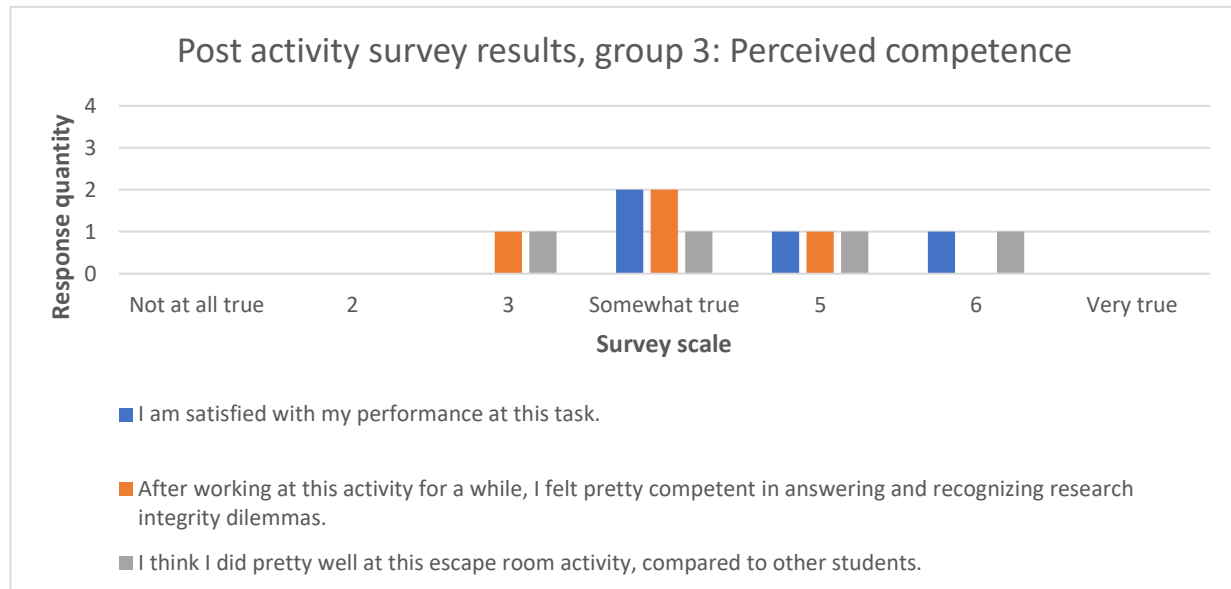


Figure 17. Survey results of all three test groups on the perceived competence category

7.3 Effort/Importance

The third motivational aspect that has been measured is effort/importance. The information on the effort/importance category will provide insights on their motivation in general and which escape room factors made them putting more effort into the escape room. Effort/importance is connected to several intrinsic motivation factors. For example, when someone thinks of the escape room activity as valuable and important, then they are most likely putting more effort in trying to complete the escape room.

According to observations, all groups did show more signs suggesting effort than signs that indicated importance towards the activity. The groups tried to participate and contribute as much as possible to the activity. Because of technical issues in the first two groups, some participants could have gotten the idea that they had put less effort into the activity because they were technically not able to. Most participants thought the activity was fun and enjoyable (interest/enjoyment) and if you really like participating in escape rooms and solving puzzles, than you want to perform well. You want to complete the rooms, else you get the feeling you failed at the activity (negative perceived competence). This fanatic behavior could be the reason for participants believing it was important to them to perform well at the activity. The first group put the least effort of all groups in discussing the dialogue dilemmas. Especially the last two dilemmas were discussed very little. Not only the first group, but all groups showed signs of fading interest towards the third and fourth dialogue dilemma. This was caused because of the missing connection between the dialogues and the rooms (discussed in the perceived competence category), but also because the decisions they made did not impact their escape room experience at all. So for the participants' experience, it did not matter at all which decision they made to answer the grey dilemmas. They found out that the decisions were not impacting the experience after the second dilemma, which created less interest and effort towards the third and fourth dialogue dilemmas. When the decisions they made influenced their escape room experience, they probably would have put more effort in the grey dilemma discussions and decision making. All groups tried their best to complete all the rooms in time, even with technical issues in group 1 causing two participants to spectate the leading PowerPoint shared in Microsoft Teams. Participants always persisted when technical issues jammed their escape room experience, but also

when the puzzles were too hard to figure out. The moderators provided hints when groups just were too stuck to proceed, but the groups never gave up and never expressed signs of wanting to quit. Most participants liked (too) challenging puzzles over easy puzzles. The third room (Doctor's office) was the most challenging, but the more challenging the better according to several participants (as long as the time limits are reasonable). All groups also communicated as well as possible in the online setting via Microsoft Teams. However, the only group that came close to dividing room tasks was the third group. When groups are not dividing tasks, completing the room in time is going to be hard. The first group for example spend more than 5 out of the 15 minutes given for the room, for solving one part of one puzzle. All four participants focused on this task, but it was possible to complete the task with two participants only.

The focus group provided information on the thoughts on the effort/importance of the participants and confirmed most observations made during the activity. The first group put a lot of effort into just participating in the first place. Two participants had either a slow device or a MacBook which was less compatible with the PowerPoint-file than a Windows device. Despite long efforts to still participate on their own device, the two participants were forced to participate through the shared screen in Microsoft Teams from the leader participant (spectating). Collaboration and communication within the rooms did almost only occur between the two participants that were actually playing on their own devices. The spectators tried to blend in and tried to make the best of the situation:

Group 1, student 1: I think the collaboration went quite well, like it was a bit difficult in the rooms due to people not having their own work and PowerPoint. So we could not see as many things as we wanted at the same time. But, like discussing the dilemmas went well.

According to the previous quote, the participants felt like the dialogue dilemma discussions went well. However, the participants did not always discuss the dilemma thoroughly and all agreed mostly rather quick on one of the options. For some participants, an external factor like group pressure could have caused the quick agreements in the dilemma discussions. A dilemma discussion example from the first group:

Group 1, student 2: Has everybody read all four options.

Group 1, student 1: Yeah, you always need a hypothesis first before you start a research.

Everyone: agrees.

Group 1, student 2: So we will go with that option?

Everyone: agrees.

The second group also encountered technical issues, but no participant had to spectate the leader participant in Microsoft Teams. Every participant tried to contribute as much as possible during the dialogues and the completion of the rooms, which led to interesting discussions. The collaboration and communication between the participants improved while progressing through the activity, but the lack of visual images (cameras) and the element of time pressure inside the dialogues and rooms limited the communication:

Group 2, student 4: Our collaboration and communication improved while we progressed through the rooms. So it is improved in my opinion, but it is still a little bit difficult because you cannot see each other.

Group 2, student 3: And then I think it also plays a role that you put into a room with three strangers and you do not really want to take the lead and tell them what to do. But at the same time, maybe that is necessary.

Group 2, student 3: I think because of the time limits, often it was difficult to let everybody say something and think about it a little bit.

The third group did not encounter technical issues at all and progressed the most fluent through the dilemmas and the rooms. At first, the participants needed to get used to the online situation. But the same as the second group, the collaboration and communication got better while progressing through the activity:

Group 3, student 4: At the beginning it was rather difficult to communicate and collaborate, but after that we discussed pretty good in my opinion.

Group 3, student 3: We grew. I think what student 4 said, we grew in our communication. In the beginning we all did the same thing and in the last room we divided tasks and shared information efficiently.

The survey outcomes on the effort/importance category show that 10/12 participants felt like they put somewhat effort or more into the completion of the escape room (see figure 18), which are rather similar to the observations and focus group outcomes. Two participants rated their effort rather low (2 and 3 out of 7), which could have been caused by the technical issues and the fading interest towards the dialogues. Even though signs of importance were hard to find during the escape room, the participants felt like it was rather important to them to do well at the escape room. This feeling of importance might originate from wanting to know more on how to conduct research responsibly and how to recognize and react towards unforeseen situations like grey dilemmas. A feeling of importance towards the activity is directly connected with two other motivational aspects, interest (section 7.1) and value (section 7.7). When someone finds contents interesting or valuable for their own development in conducting research responsibly, that content is probably important to them. So when the research integrity contents of the escape room are interesting, valuable and recognizable, the participants are able to feel like it is important to them to do well at the escape room activity. According to the results in the *interest/enjoyment* category, the interest towards the escape room contents under the participants was rather high. The importance interpretations made based on the survey results are not confirmed by the participants, as the survey results were not discussed afterwards.

Overall, the results indicate signs of effort and importance towards completion of the escape room activity. The participants put the most effort in the completion of the rooms. The missing link between the dialogues and the rooms caused a fading interest towards the last two dialogues, with less effort as a result compared to the effort given inside the rooms. Therefore, the effort that the participants showed, positively indicates motivation towards the completion of the escape room activity.

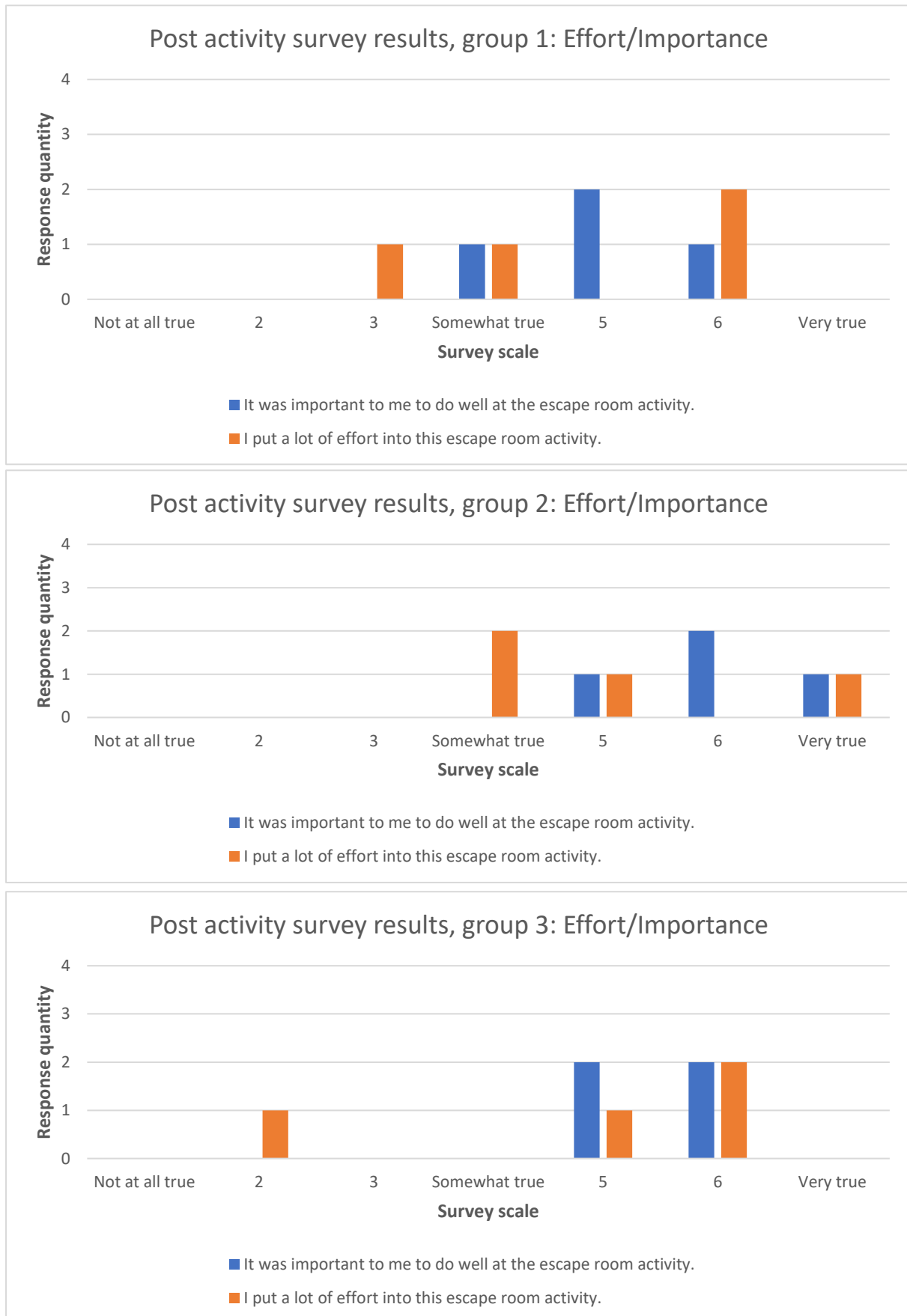


Figure 18. Survey results of all three test groups on the effort/importance category

7.4 Pressure/Tension

Pressure/tension is a negative predicator category towards intrinsic motivation. Pressure is an extrinsic factor that is able to let people make other decisions than they would make without this pressure.

Observations did not really show much signs on pressure, tension or nervousness. Most groups were aware of the time limit, but it did not make them rush the puzzle route inside the rooms. In some dilemmas the impact of time pressure was noticeable. The leading participant in the second group for example, clicked on an answer before a consensus was reached on deciding what to do with a grey dilemma because the discussion time almost ran out. So observable signs of pressure were mostly caused because of the timers inside the rooms and dilemma discussions. Another possible pressure factor that did not show any signs during observations, is peer pressure. The participants are working in groups in an online environment where they cannot see each other on a camera (because the escape room is played in full-screen) and the group members of the first and second group did not know each other at forehand. This might have caused difficulty in disagreeing with the group when participants thought of other answers. These pressure creating factors are confirmed by all groups in the focus groups, which suggest a negative impact on intrinsic motivation of the participants:

Group 1, student 2: You kind of due to the time pressure, follow what you already thought.

Group 1, student 4: I think with some of the answers we gave as a group, I might have not completely agreed with them.

Group 2, student 2: I think that especially time pressure plays a role often because most people know actually what is the right thing to do, but time pressure is the problem.

Group 2, student 3: And then the second time the same question came again. I did not even check all the answers, partly because of the time pressure and partly because we discussed that before. So in this case, I think the time pressure played a role in that way.

Group 2, student 3: I think it also plays a role that you put into a room with three strangers and you do not really want to take the lead and tell them what to do.

Group 2, student 3: I think a minus point would be that we could not see each other when we were talking to each other.

Group 2, student 4: I think the communication improved, but it is still a little bit difficult because you cannot see each other.

Group 3, student 2: It did not feel like we worked together really well. I also could not see my group which gave me a distant feeling towards them.

Group 3, student 2: Because it is a cooperative activity, you mostly provide answers that the other group members want to hear.

Something to mention is that some participants had two screens, one screen with the PowerPoint and one with the Microsoft Teams meeting. This caused these participants to still see their teammates while also participating in full screen mode. The second screen could have decreased the distant feeling and pressure towards other participants:

Group 3, student 4: I was glad I had two screens, because then I just saw you guys.

But the pressure created by time was not always a bad thing or feeling. The participants recognized the feeling of pressure and most of them were able to link this type of pressure to their own research

practices and assignment deadlines, which created more meaning and value towards the escape room activity and a slightly positive influence on their intrinsic motivation towards the escape room:

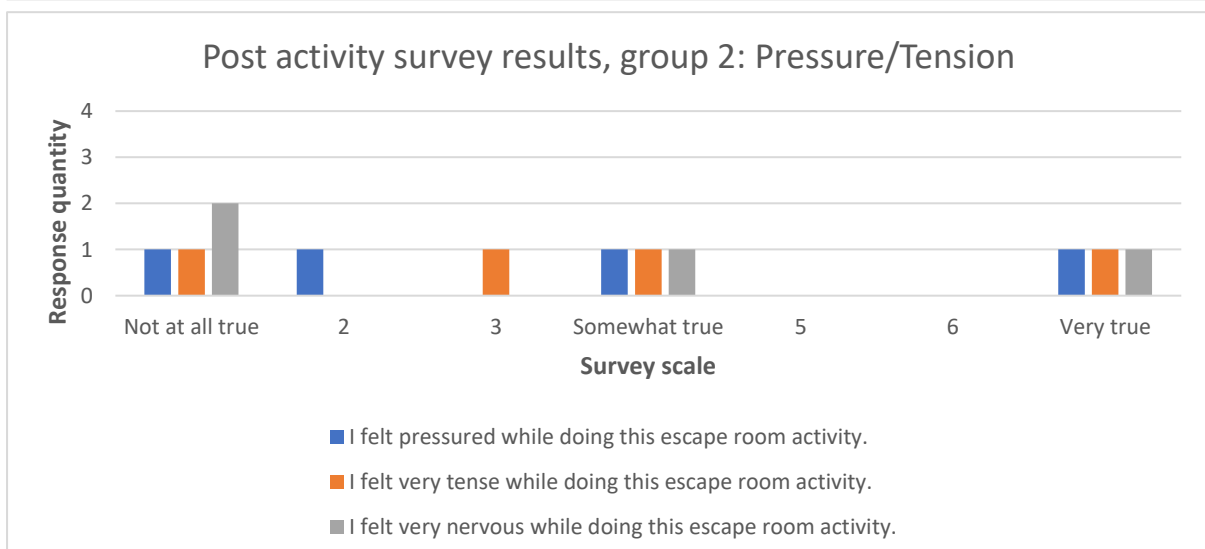
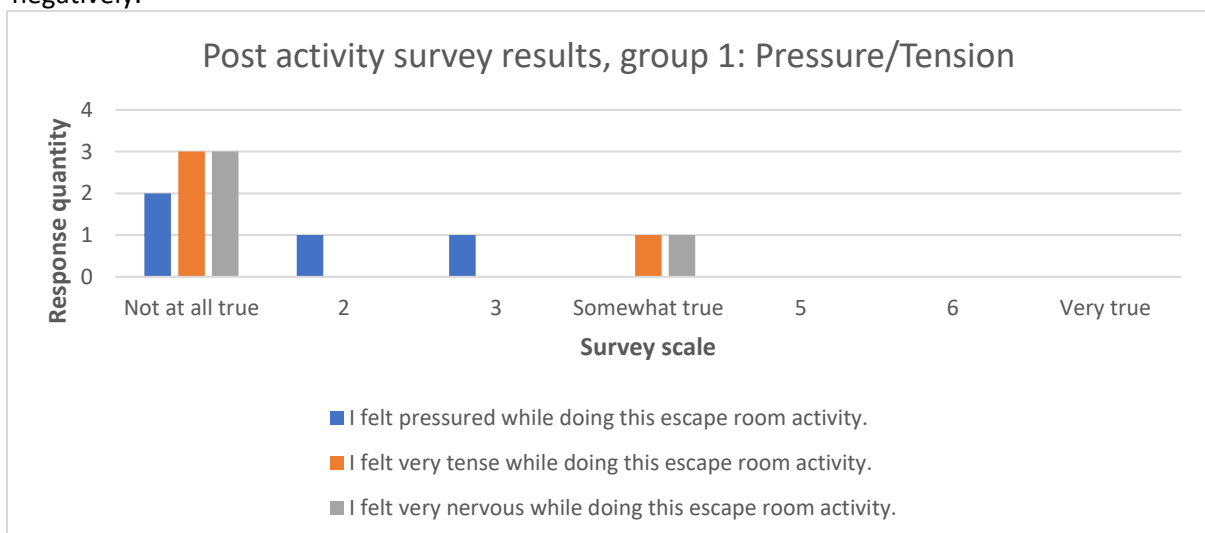
Group 1, student 1: The part with the 'oh we have a deadline' and then you have to do research, that was recognizable to me. Time pressure of all the people around the lab drive me to make certain decisions.

Group 2, student 2: I find it a bit difficult to have empathy for the fictional characters, but I do know that there is often experiments, especially done by students, with time pressure. I do not feel that time pressure so much.

Group 3, student 3: I never submitted an assignment too late because of time pressure, but I am able to understand that other people recognize the deadline situation.

The survey results on the pressure/tension category show that feelings of pressure, tension and nervousness were present for several participants (see figure 19). A few participants even felt no pressure, tension and nervous feelings at all, especially several participants of the first group. Then in the second group one participant even scored very true on all three the feelings.

The online situation (with no vision) combined with a formed group of strangers caused most of the peer pressure, tensed and nervous feelings. Both peer pressure and time pressure played significant roles in the escape room experience, influencing intrinsic motivation both positively and negatively.



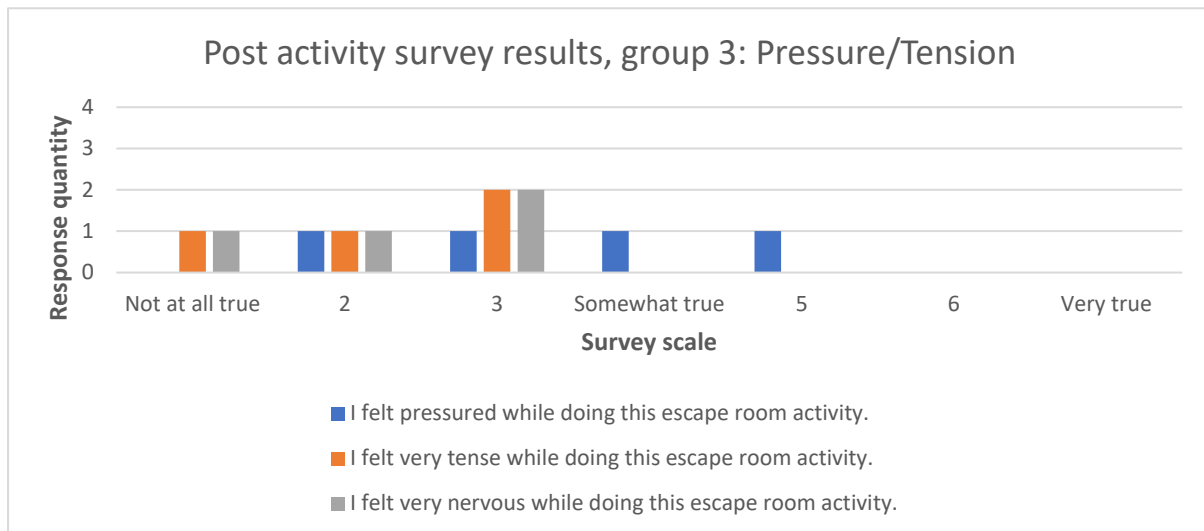


Figure 19. Survey results of all three test groups on the pressure/tension category

7.5 Perceived choice

The perceived choice category in the data-collection represents the need of autonomy that is needed to achieve intrinsic motivation. Perceived choice focuses on the freedom of choice by participating to the escape room. In order to feel autonomous, a participants needs to feel freedom of choice towards their own actions and insights. A participant therefore needs to have the possibility to influence and control their own actions whenever they want to. The escape room tried to create a feeling of autonomy by using an open-puzzle structure, providing multiple answer options to the dialogue dilemmas, creating the possibility to reverse dialogue slides to read and analyze them at the participants' own pace, and the ability to choose between two puzzle routes which both lead to the final code of the room. The observations and focus group measured signs of autonomy during the escape room, whereas the survey focusses on the participants' freedom of choice for participating in the first place.

The observations regarding perceived choice, focused on signs that participants did or did not agree with the given options within the dialogue dilemmas and their freedom inside the rooms. The most noteworthy observations about the perceived choice category, were the observations from the third group. Almost every grey dilemma the participants encountered in the dialogues caused one or more participants to note that the answer they would like to give, was not one of the four available options that were given. The other groups showed no observable signs of (a lack) of autonomy, but that does not exclude that there are none at all.

Regarding the focus group outcomes, the third group confirms the perceived choice observations. They would have liked to formulate own ideas and answers rather than picking one of the four given options. This limited their feeling of autonomy in the dialogues and negatively influenced their intrinsic motivation:

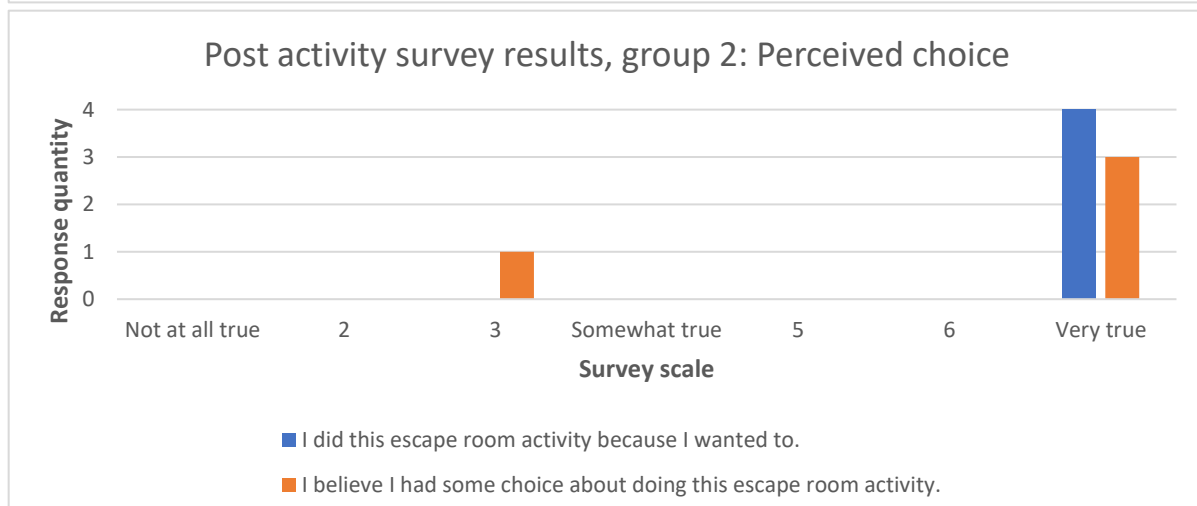
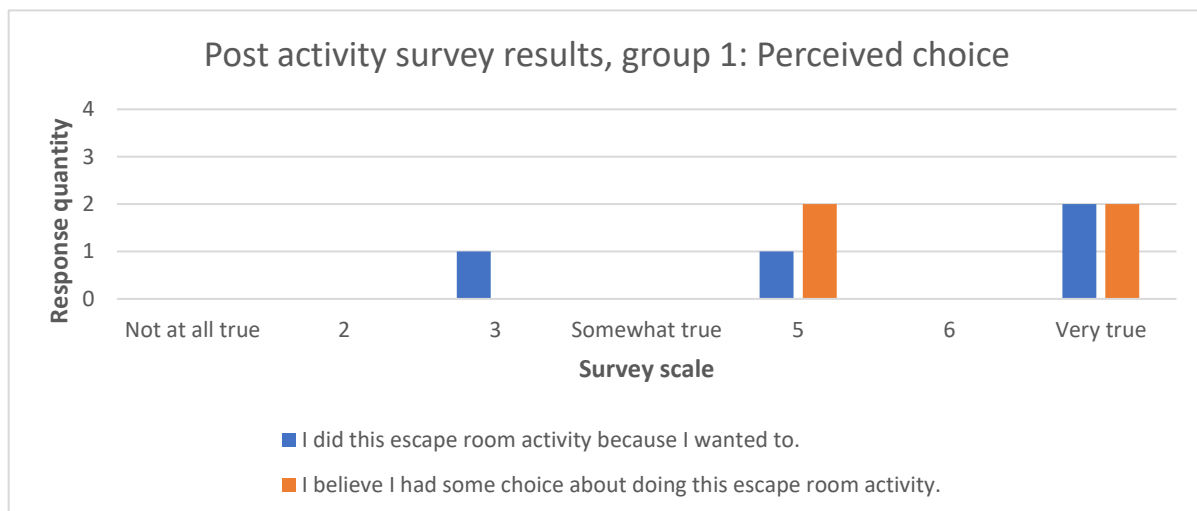
Group 3, student 3: I would not choose any of the options that where provided when answering the dilemmas.

Group 3, student 3: We thought a lot of times like 'no we would not do all of those answers and we prefer a different option'.

Group 3, student 4: Yes, we did not find our perfect answer in between those given options.

The survey statements on perceived choice were mainly focused on the motives of participating in the first place (see figure 20). To find participants, master students needed to participate voluntarily. The first two groups participated voluntarily, but the third group was selected from a university class to participate (so not completely voluntarily). When looking at the survey statements in the perceived choice category, a maximum score was expected for all participants in the first two groups (because they volunteered). However, the survey scores show some remarkable outcomes. The third group showed a variation in survey outcomes as expected, because they did not all volunteer. However, the survey results of the first group also showed lower scores than a 7/7. This would indicate that two out of the four participants of group 1 somehow did not participate voluntarily (because they wanted to). The second statement states that the participants believed they had some choice about doing this escape room activity. It might be a possibility that the participants misread the statement, linking it to the available choices inside the rooms. When looking to the second group for example, one participant scored a 7/7 on wanting to do the escape room, but scored a 3/7 on believing he or she had a choice about participating. It is quite remarkable to see these varying results coming from volunteering participants.

According to the results, most participants felt like they were able to choose options that were in line with their own thoughts and were able to navigate freely through the escape room. Only the third group was not feeling autonomous regarding the available options given to solve the grey dilemmas, discussed in the dialogues. Overall, most of the elements used to create autonomy were influencing the participants' feeling of autonomy positively, only for the third group the dialogue dilemma design influenced their intrinsic motivation negatively.



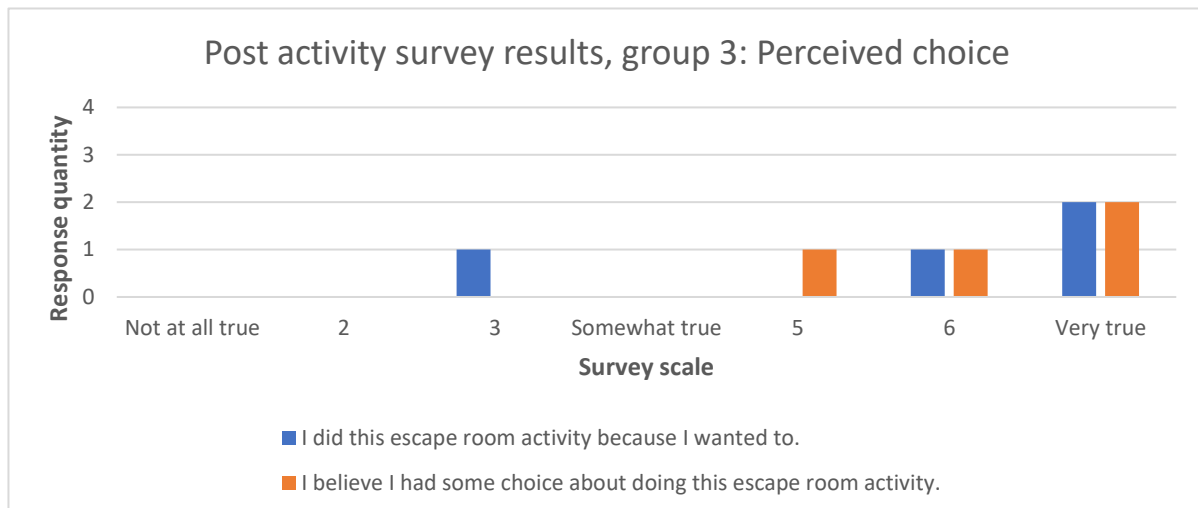


Figure 20. Survey results of all three test groups on the perceived choice category

7.6 Relatedness

The relatedness category measures the relation between the participants during the escape room activity. Relatedness is one of the three cognitions that someone needs to feel optimally to be intrinsically motivated towards an activity or content.

Observing behavior relative to relatedness indicates the effect on the participants' intrinsic motivation. Behavior specifically related to the relatedness category were hard to identify. Based on communication, almost none of the participants showed inactivity or long periods of silence. All groups tried their hardest to work as a team and to discuss the grey dilemmas until a consensus was reached. Only the first group had two participants who were rather silent throughout the activity, which could be linked to a distant feeling towards the other group members. These were the two spectating participants with technical issues, which could also have caused the inactivity.

Based on the focus group outcomes, more information on the relation between the participants aroused. More specific, how the participants felt towards their group members. The first group did pointed out that they felt distant towards their group members, even though they did not know each other before participation. However, the second and third group did describe the relationship between their group members specifically. Both groups mentioned that the online setting in Microsoft Teams did not make them feel comfortable when communicating. Because of the lack of visual images of group members and the fact that they did not know each other at beforehand, the second group struggled at first to match the voice they heard to the correct person (with the third group not having that issue):

Group 2, student 2: While answering a dilemma, I thought we had already decided. I was confused with their voices.

Group 2, student 3: It also plays a role that you got put into a room with three strangers and you do not really want to take the lead and tell them what to do.

Group 2, student 4: Communicating is difficult because you cannot see each other, but it improved while progressing through the activity.

Group 3, student 3: Fortunately, we knew each other because when we were strangers to each other I think it would be less fun and maybe even annoying.

The third group already knew each other from the same university course they followed at the time, so unfamiliarity did not influence the relation between the group members. What did influence their feeling of relatedness, was the lack of visual images. All participants in the third group spoke of a distant feeling towards their group members because of not being able to see them during the escape room. Some did not feel this distance towards the others, but that was mainly caused by a second screen where the participant then still was able to see the other group members while also completing the escape room in full screen on the first screen:

Group 3, student 1: It did not feel like we communicated and worked that much together. Also, I could not see my teammates which gave me a distant feeling towards them.

Group 3, student 1: We are lucky with the distinctive voices that we have as a group. For example, you will recognize the voice of student 3 everywhere! Else it would have been very unclear who was talking without seeing them on camera.

Group 3, student 3: I understand that we have to work at a distance, but I do not know if it is ideal to wait at each other most of the time.

Group 3, student 4: I was glad I had two screens which provided me with your images while playing the game.

One participant from the third group stated he or she was glad they knew each other because when they were strangers, it would be less fun and maybe even annoying. These results of the relatedness category are examples that indicate the many connections that all seven motivational aspects have with each other.

The relatedness statements used in the survey, measure the relationship between the participants (see figure 21). The participants' feeling of distance and trustworthiness towards other group members were measured statements. All groups thought they could trust their teammates during the activity to rather high extends (with a minimum score of 5/7), which are rather high scores especially for the groups that consisted of participants that were unfamiliar to each other. The survey scores on the statement representing a feeling of distance towards group members, confirm the observations and focus group outcomes of the second group. Three out of four participants felt somewhat or more distant towards their group members. The same scores were expected from the third group, but these scores are very low with a maximum score of 2 out of 7. These scores were partially unexpected, because the participants did spoke out of their distant feeling in the focus group. The low scores might originate from the second screen that most participants had, which enabled them to still see the other group members.

Summarizing, when the feeling of relatedness between participants is negatively impacted by the design choices of an activity, this could cause distant feelings towards other group members and could even influence the participants' enjoyment negatively. The feeling of relatedness that is created by the activity directly influences the intrinsic motivation that the participants have towards the completion of the escape room activity.

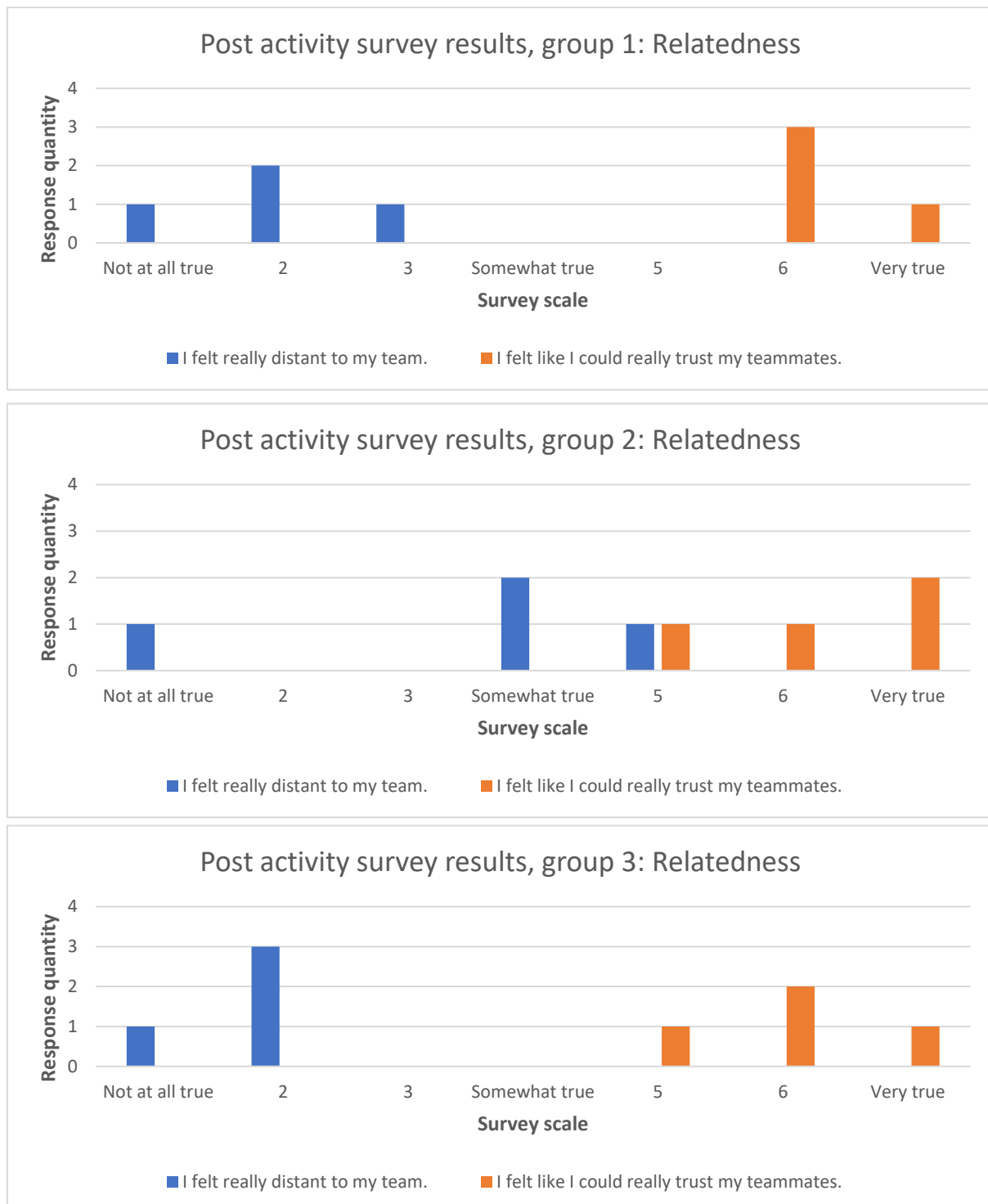


Figure 21. Survey results of all three test groups on the relatedness category

7.7 Value/Usefulness

The last motivational aspect that is measured in this research is the value/usefulness category. When someone values an activity, goal internalization and self-regulation is influenced positively. Creating a valuable activity for participants also contributes to the generation of meaningful play, which arouses motivation to take part in that specific activity.

This certain motivational aspect was not observable during the activity, but the focus groups generated interesting information on the value/usefulness category. When a participant is not involved in an educational environment, the use of an educational escape room will most likely not interest the participant. When the content of the educational escape room did nothing for the participant and did not grant him/her the feeling to have performed well or to be more confident about making decisions when encountering grey dilemmas (competence), the value/usefulness scores will most likely be rather low. Most participants thought the escape room activity was valuable and useful to them because they were able to relate the research integrity contents together with the recognizable storyline and time pressure, to their own research practices. These participants might have felt that the activity could contribute to their development process in being able to conduct research responsibly (RCR) and being able to react to unforeseen situations (like grey dilemmas). This last note is an interpretation that is not confirmed by the participants, but the participants of all groups did underpin their believes about the activity being of (some) value to them:

Group 1, student 1: The part with the 'oh we have a deadline' and then you just have to do research, that was recognizable to me so that you have, like time pressure of all the people around the lab.

Group 2, student 2: Well I do a lot of lab work. And the part where the scale was not working, I actually experienced that myself. So I feel for him (fictional character), but it was a lot of drama and I do not think that in real life it is that bad.

Group 2, student 3: Especially the specific dilemma about the director reaching out to you with some pressure was very recognizable to me. Yeah, I guess the other ones about the teamwork are also good ones.

Group 3, student 1: It is very typical from a supervisor to say 'it has to be more!' or 'again, it is not good enough!'

Group 3, student 2: I recognized some parts of the activity in my own research practice, but the available options for the dilemmas are options I would not have chosen in that situation.

Group 3, student 3: Everyone recognizes those research dilemma we came across in the activity. For example, someone who just does not cooperate in the group.

Many participants recognized several research elements from their own research practices, but few participants thought it was hard to feel empathy for the fictional students in some of the research situations. Some aspects of the narrative just did not seem very logical for a few participants:

Group 2, student 1: I find it a bit difficult to have empathy, but I do know that there is often experiments with time pressure. But with 10 or 20 patients and all the time in the world to collect data, it is not going to be enough for some serious thesis or research anyway. So basically, the research is not serious from the beginning.

Regarding the usefulness of the design, the third group came up with (lesson) inspirations and other ideas for uses of the escape room, caused by the escape room design and experience:

Group 3, student 2: I think this activity could probably suit bachelor students as well, besides the targeted master students. The difference is not that high between master and bachelor students in some areas.

Group 3, student 3: It might even be a suitable activity for an introduction day or week!

Group 3, student 4: I also think that the activity is suitable as an orientation for freshmen.

Even though the third group was enthusiastic about the different options and ideas of using the escape room activity, participants of the second group thought the activity was not that suitable as intended for (master) students:

Group 2, student 3: I think for philosophy of science master students, it might be a little bit boring because you discuss these kind of issues extensively. It feels a bit childish to be presented in such kind of game I think. I think it is a good tool, but maybe for some people it is just a bit simple.

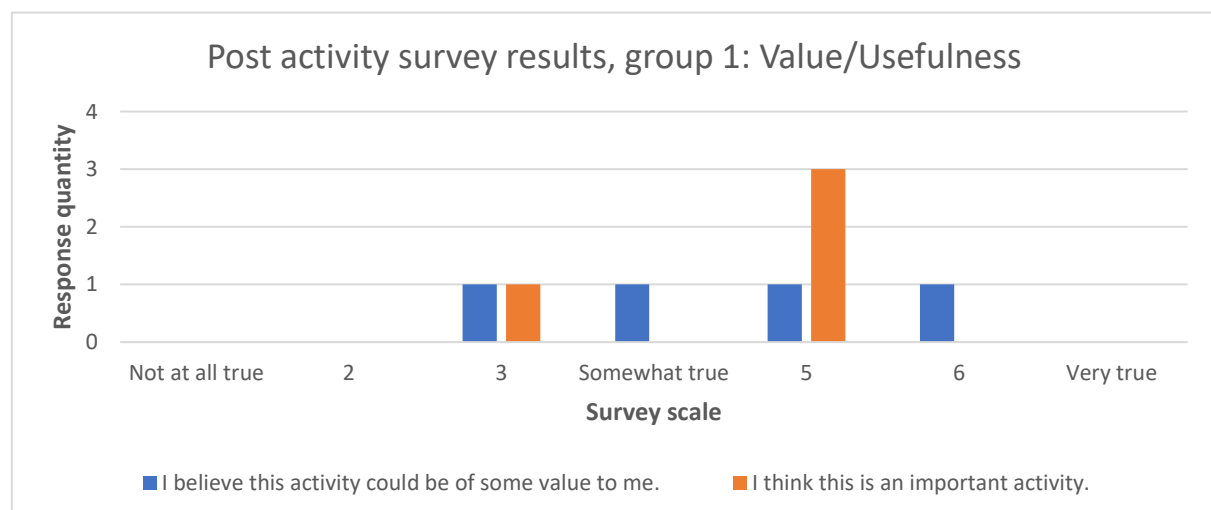
Group 2, student 2: I agree as well. I mean, some master students have gone through the bachelor in three years and some have gone through the bachelor in five years. You don't have to study philosophy to learn about this stuff.

The survey results confirm the statements that the participants made during the focus group (see figure 22). Several participants are valuing the escape room activity somewhat or more (a score of 4/7 or higher), which are caused by the recognizable storyline, time pressure and research contents. Only a few participants thought the escape room activity was almost of no value to them, probably caused by a lack of empathy towards the fictional characters and some unrealistic narrative situations (like a 10 minute deadline for the data analysis).

The third survey question regarding the value/usefulness category of the survey was opened: *I think that doing this escape room activity is useful for: ...*

The question generated interesting information on the usefulness category from most participants, but two participants did not feel the urgency to provide an answer to the question. Most returning answers to the question are: *It was a good activity for introducing research integrity; Learning more about grey dilemmas in research integrity (by discussing them with others); Lesson activities/approaches.* These answers also confirm the statements in made in the focus groups.

In short, the escape room activity was somewhat valuable to several participants due to the recognizable storyline, research contents (on integrity) and time pressure. A few participants were not able to relate to some unrealistic situations in the narrative, like a 15 minute timer to complete a whole data-set. These timers were of course fictional and represented several hours, days or weeks in real-life definitions. Most participants also felt like the activity was useful to them or to educational institutes, to use as lesson method to introduce research integrity and to learn more about grey dilemmas and unforeseen situations while conducting research. These mostly positive statements of value and usefulness indicate that the participants felt like the activity had meaning to them, contributing to a state of meaningful play for the participants. This functions as a motivator for the participants to complete the escape room activity together with the other six motivation aspects.



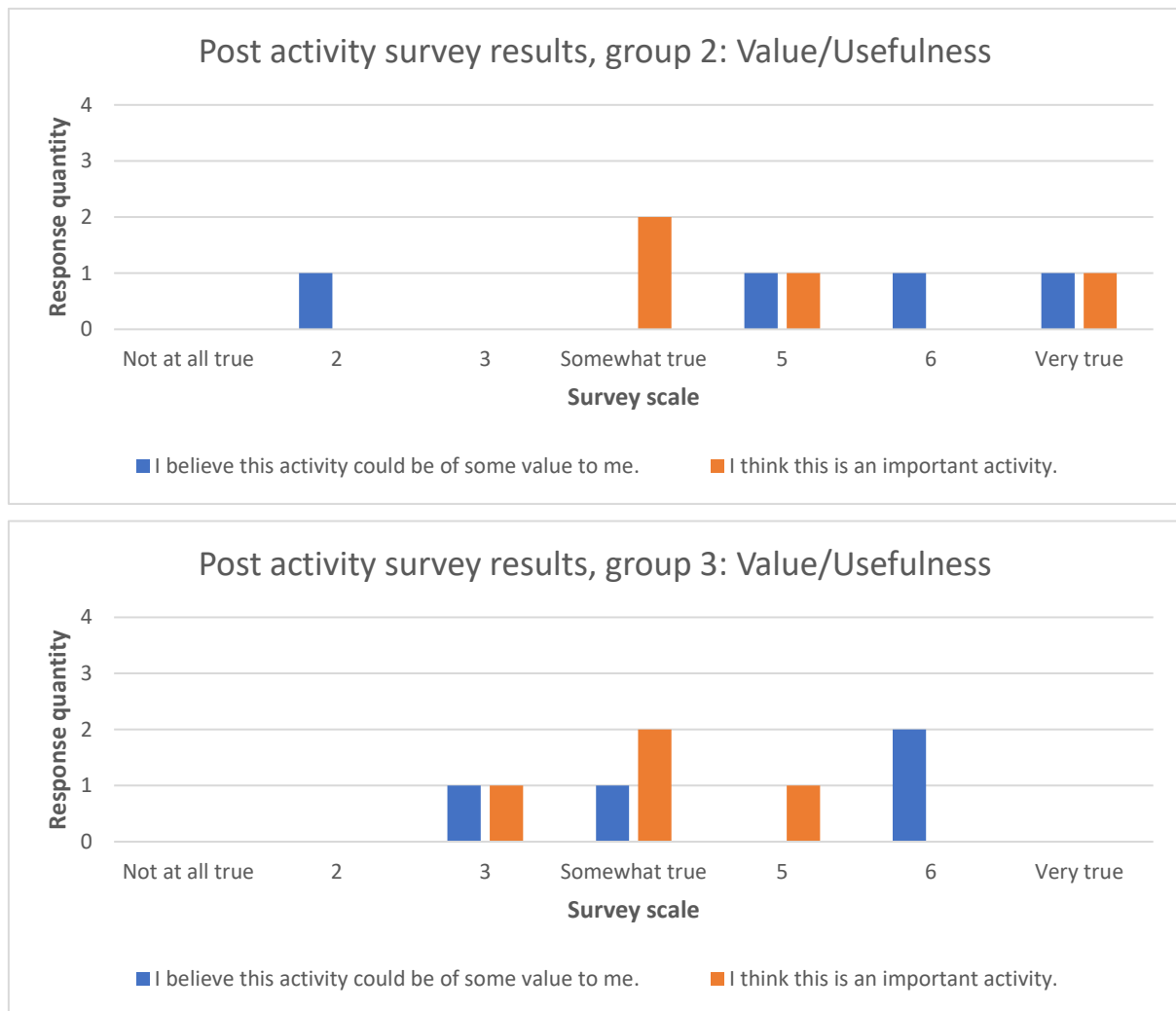


Figure 22. Survey results of all three test groups on the value/usefulness category

8 Discussion

8.1 Qualitative result interpretations

The goal of this study was to arouse master student interest and intrinsic motivation towards research integrity contents. The results of the test groups look favorable towards positively influencing the participants' intrinsic motivation at first, based on the participants' *interest/enjoyment, perceived competence, effort/importance, pressure/tension, perceived choice (autonomy), relatedness and value/usefulness* towards the activity. The participants' *interest/enjoyment, perceived competence and perceived choice* are directly related towards their intrinsic motivation. The other motivational aspects do have a more indirect relation towards intrinsic motivation, but are still relevant indicators related to motivation in general.

All participants thought the escape room activity on research integrity was fun (they especially enjoyed solving the puzzles and completing the rooms) and interesting towards the research integrity contents discussed in the grey dialogue dilemmas. The rooms were challenging and the participants thought they were fun and engaging, even with the time pressure. Participants also felt rather competent towards their escape room activity performance. This feeling was mainly caused because the participants thought the activity enabled and learned them how to recognize and react to unforeseen, complex and grey situations and dilemmas better in their own research

practices. All participants put reasonable effort into the completion the rooms, even when technical flaws caused a lot of frustration. Outcomes on the pressure/tension category indicated that the participants liked the element time pressure because of two reasons. The time pressure caused the rooms to be an actual challenge, which motivated them even more to complete the room (which indicates an extrinsic motivator), and time pressure is relatable and recognizable towards their own research contexts during their studies. Not only was the element of time pressure a familiar element, also the storyline/narrative and research integrity contents were recognizable for the participants. These elements creating meaning for the participants towards the activity, increased their feeling of value, interest and importance towards the activity as well. The participants' feelings of perceived choice (autonomy) were rather high for all groups when they chose to participate to the escape room activity (before participation). The first two groups volunteered, which should indicate intrinsic motivators driving the participants to volunteer. The third group was selected from an obligated class, which resulted in less perceived choice results before participation. According to the participants' feeling of autonomy (freedom of choice) during the activity, most participants thought they had enough options to choose from to move freely (to some extent) inside the escape room activity. The relatedness feelings inside the groups were mostly positive. Focus group and survey results suggested that the participants were able to trust each other and observations showed communication in every room and elaborate discussions (varying per group and dialogue). This should indicate a good relation between the group members of all groups, creating a safe environment for the participants to talk, discuss and act feeling comfortable. This environment positively influences the stimulation of intrinsic motivation indirectly. Outcomes of the last motivational category, value/usefulness, showed why participants valued the escape room activity and what uses they thought the activity could have for educational practices. The participants thought the activity was mostly useful for the introduction of research integrity, learning about grey dilemmas in research integrity by discussing them with others and inspiring participants for lesson activities.

Next to the positive effects that several escape room design elements had on the participants' intrinsic motivation, some elements also negatively affected the escape room experience and therefore also influenced the participants' intrinsic motivation negatively or not at all (not as intended). Firstly, in order to create cohesion between the dialogues and the rooms, the grey dilemmas discussed in the dialogues needed to connect with the research contents inside the next room. However, the link between the dialogues and the rooms was missing and while completing the rooms, participants often lost the goal of the escape room activity in general: to conduct research as integer as possible. Most participants were just trying to solve the puzzles and complete the tasks in time, rather than thinking about what puzzle route they needed to follow regarding integrity differences. The missing connection between the rooms and the dialogues together with not enough clarity and direction inside the rooms, led to a rather negative student flow-state. Students got frustrated when they did not exactly know what to do inside the rooms, also because they had a time limit for every grey dilemma and room. This suggests that the element of time pressure also influenced the participants' escape room experience negatively, and with that negatively stimulating their intrinsic motivation towards the activity. This result regarding pressure was expected, as pressure/tension is a negative predicator category towards intrinsic motivation. Furthermore, the participants' stated that they had learned to recognize and react to grey dilemmas and unforeseen situations while conducting research better because of the escape room activity. All the groups learned that it is always an option to discuss integrity issues with a supervisor or fellow students, rather than struggling to solve the problems alone. However, the participants also stated that they did not learn how to deal with the grey dilemmas on their own. This was caused by the lack of

confirmation provided by the cooling-down period. The focus group nature of the cooling-down period focused more on the interests of the researchers, rather than the participants' interests. Also, not every aspect of the escape room was discussed, causing little confirmation on the decision making of the participants. This design choice caused too little perceived competence towards the answering of the grey dilemmas, lacking the intended positive effect on their intrinsic motivation. Regarding the effort/importance and the relatedness category, two technical factors influenced the participants' escape room experience negatively caused by the online setting. The first factor is the incompatibility of MacBook devices with add-ins in Microsoft PowerPoint. This caused the most issues for the first group, resulting in two participants not being able to play the escape room on their own. They needed to spectate the leading PowerPoint that was shared in the Microsoft Teams meeting in order to participate to some extent. This negatively influenced the flow-state of the two spectating participants, but it was also uncomfortable for the still playing participants because they needed to take the spectators into account by not going too fast through the slides. This limited their freedom inside the rooms (autonomy) and it could have also limited their effort in discussions and solving puzzles. The second technical factor is the absence of visual images of group members. The participants did turn on their cameras for research purposes, but were invisible to the participants because the escape room in PowerPoint needed to be played in full-screen. This caused the participants to not being able to see each other. This was especially an issue with the first two groups because they did not know each other before participation, but also the third group spoke out of the uncomfortable online setting. This factor caused some participants to feel rather distant towards their group members, negatively influencing the relationship (relatedness) and indirectly the intrinsic motivation to contribute to discussions and solving puzzles.

8.2 Study approach and methods

The approach used in this study was based on an escape room design framework by Clarke et al. (2017) and escape room design guidelines stated by Eukel & Morrell (2020) (see section 4.1, figure 3). Not every aspect of escape room design framework by Clarke et al. (2017) was used in the same order. In order to design the escape room, first the theme was discussed and specified: a positive focus on research integrity (RCR) rather than a negative focus (FFP) to create awareness on unforeseen situations and grey dilemmas in research practices. The theme specification was in combination with determining the participants (master students), the objectives (learning and performance goals) and the necessary equipment (online setting, Microsoft Teams and Microsoft PowerPoint). After these four aspects were specified the designers chose to design and create the escape room activity, before creating the methods to collect data. After the activity and data analysis, this proved to be a questionable decision. When the data collecting methods would have been created before the escape room was designed, the design choices would have been more well-thought, would have shown much more cohesion with the methods and would have generated more valid and reliable results. After creating the methods, the evaluation part was split up in a pilot evaluation and a test round evaluation. The pilot was needed to test the functionality of the escape room design, and the test rounds were needed to generate results to the used methods. The re-evaluation phase provided improvements and recommendations of this design study for other related studies. The framework mostly proved to be effective, but the methods designing phase and creating the escape room should be switched around. This framework decision creates an urge to reflect the used methods, to eventually connect the results of this study to the most valid and reliable conclusion possible.

In the method section (chapter 4), three data-collecting methods are elaborately discussed with the used instruments available in the appendices (A, B, C and D). These methods tried to ensure

data triangulation, to significantly increase validity and reliability of the results. However, there are some irregularities regarding the methods that need to be addressed.

At first, the methods are designed to generate the same type of answers towards the intended motivational aspects that are measured in this study: *interest/enjoyment, perceived competence, effort/importance, pressure/tension, perceived choice (autonomy), relatedness and value/usefulness*. All methods succeeded mostly in providing the same type of answers and results from the motivational aspects (validity), except the data collection on the *perceived choice* category and the *relatedness* category. The perceived choice category measured the participants' feeling of autonomy before participating and during participation. However, the observation and focus group instrument only focused on the feeling of autonomy *during* the escape room activity and the survey statements focused on the feeling of autonomy while choosing to participate in the first place. This negatively influences the validity and reliability of the results based on autonomy. This irregularity could have been solved rather simply, by discussing their freedom of choice to participate in the focus group and by adding autonomy related statements about their participation on the escape room activity in the post activity survey. These insights aroused too late to adjust the methods in the test rounds. Next is the *relatedness* category, especially the use and adjustment of the relatedness survey statements in this study (based on the IMI). The IMI was used to create the post activity survey on intrinsic motivation, but the scale indicates that the latest category was not yet validated. The latest category was *relatedness*, implying that these statements were not validated compared to the other categories (which are validated). Lastly, the three data collecting methods are all created and adjusted based on the IMI. The survey questions are adjusted and formed to escape room activity statements and the focus group questions and observation form are focused on finding information about the seven aspects of intrinsic motivation used in the IMI. The observations and focus group results then are confirmed (or not) by comparing them to the survey results. In short, the IMI is validated for 6/7 motivational aspects, but the altered data collecting methods used in this study are not validated professionally.

9 Conclusion

This design study served as a kick off for the escape room project and valuable insights will be used in future studies and designs regarding the project. This study was aimed to answer the following research question: *How does an educational escape room on research integrity affect the motivation of master students towards research integrity?*

The answer to the research question was generated from the insights and perspectives provided by answering the following sub-questions:

1. How are master students motivated during the escape room activity on RCR?
2. How do escape room elements affect the intrinsic motivation of master students?

According to the study results, regarding the first sub-question, the participants thought at first that it would be fun to participate in an online escape room. Before the escape room started, there was no specific interest in the content (research integrity) of the escape room. Some of the participants just discovered the research integrity theme when they started the experience. Most participants volunteered to participate, which indicates intrinsic motivation towards participation. During the escape room activity, participants enjoyed the puzzles, solving riddles and completing the rooms the most. The participants thought that the dialogues and grey dilemmas were interesting, but lacked cohesion with the rooms. This interest faded a little after two dialogues, because the decisions the participants made had no consequences to their escape room experience. This indicates that the

participants at first started the activity intrinsically motivated to try and complete the escape room as fast as possible, but their intrinsic motivation was not as strong in the last two dialogues and rooms compared to the first two dialogues and rooms. Not solving a room created a feeling of failure and frustration for especially the first two groups of participants (who did only solve one or two rooms). Escape rooms generally do not have the chance to redo the experience, excluding the concept of recoverable loss. Their feeling of competence towards the activity was influenced negatively or not as intended, because of the unsolved rooms and lack of confirmation on the decisions they made to the dialogue dilemmas. The last factor that made several participants' intrinsic motivation fade as the escape room experience progressed, was the online communication setting which affected their relatedness negatively. Several participants felt distant towards their group members because they were not able to see each other on camera and the participants of the first two groups were not familiar to each other before participation. The escape room design did not ideally fulfill the needs that are necessary to create an intrinsic motivation stimulating environment, as the competence, autonomy and relatedness requirements were not stimulated optimally by the made design choices and used escape room elements. As the escape room progressed, multiple elements affected the participants' motivation to continue and provide as much enjoyment, interest and effort as they did at the start of the activity. That is where the sub-question 2 is being answered.

Regarding the second sub-question, the escape room elements that affected the participants' intrinsic motivation the most are: *the introduction/instruction, (a lack of) direction inside the rooms, storyline/narrative/theme, time pressure, (a lack of) direct feedback/reflection, online collaboration, increasing puzzle difficulty per room (increasing flow of complexity) and (the lack of) generating own answers to grey dilemmas*. The design tried to link as much as possible to the participants' own research practices, which had the desired effect. Participants recognized time pressure from deadlines and the storyline/narrative created more meaning, value and importance towards completion of the activity, which stimulated motivation, meaningful play and goal internalization towards the activity and learning about the research integrity. However, the element of time pressure also negatively influenced participants' grey dilemma decision making. The timers caused several participants to rush their decision making and not discussing the dilemmas thoroughly with the group. The next element, the lack of direct feedback and reflection on grey dilemma decisions (especially during the focus group in the cooling-down period), negatively impacted the participants' perceived competence towards answering grey dilemmas on their own. However, participants did learn that it is always a possibility to discuss integrity issues or unforeseen situations with a supervisor or fellow students. They learned how to recognize and partly deal with grey dilemmas (by discussing them with others), but not how to solve them on their own. The introduction was created to initiate meaningful play and help to make the experience as smooth as possible, linking the escape room goals to the activity: conducting research as integer and responsible as possible (RCR). The introduction had not the desired effect, because all groups were near to clueless what to do in the first room. The introduction proved not to be a clear enough instruction towards the escape room activity. Also, despite improving the direction of the puzzle routes inside the rooms after the pilot evaluation, the direction of the puzzle routes inside rooms was still not clear enough for all three groups. This caused the groups to lose focus on the research integrity theme inside the rooms, only focusing on completing the puzzles within the time limit. This caused the participants to not perceive competence on choosing the most integer puzzle route, because most of them did not even notice a difference in integrity between the two routes. The increased puzzle difficulty per room had both positive and negative motivational effects. The increasing difficulty made it for several participants enjoyable and challenging, but for some participants too challenging. Especially the third room (hardest room) proved too hard to solve in 15 minutes, with none of the groups solving the doctor's

office in time. Collaborating and communicating online in groups of strangers (first two groups) created a distant feeling towards the other group members, which affected relatedness for some participants. Also, a lot of participants endured technical flaws which caused frustration and limited the experience and enjoyment for them. Lastly, the design tried to create as much freedom of choice as possible for the participants, but some participants rather formulated their own answers than having to choose from four given options. Especially the participants from the third group would have liked to give a different answer than the four options the design provided for a dilemma. Other groups might have not felt completely free to move around the room because of technical issues or spectators with non-working PowerPoints, influencing autonomy negatively during the escape room activity. The information gathered from the results provided insights to generate an answer to the research question.

In conclusion, the current design is not able to fulfill all needs for intrinsic motivation, but the potential is certainly there. Besides the points of improvement, the escape room activity highlights grey dilemmas and unforeseen situations while conducting research. All participants had to think about the integrity issues and discussed them within their groups. Almost every group held elaborate and interesting discussions about different kinds of options to the dilemmas, which confirmed the fact that there is not just only one good answer to such a dilemma. Some participants even came up with own answers, next to the given options. Even though some participants already made their own choices regarding (research) integrity, the escape room activity showed the participants that not every scenario in conducting research has black and white answers or decisions to make and made them more aware of this fact. Participants mainly were engaged and had fun while solving puzzles and escaping the rooms, and were more interested towards the research integrity contents while discussing the grey dialogue dilemmas. The mixed outcomes of this design study indicate potential for the educational escape room as a motivational tool on research integrity for research master students when designed optimally, which needs further research to achieve.

10 Further research

Despite the potential that an educational escape room has to generate interest towards research integrity under master students, it is not certain that an escape room is the most efficient tool for providing the desired learning outcomes regarding empowerment, motivation and awareness. For engagement, an escape room proved to be a viable educational tool. But for learning more about research integrity, grey dilemmas and unforeseen situations in research practices, other educational tools might show better results. The escape room element that clashes with many learning goals is the element of *time pressure*. In this design, time pressure had positive and negative effects on the participants' intrinsic motivation and learning process during the activity. When removing the element of time, discussions could be more elaborate and decisions might be able to be more well-thought. But is an escape room without the element of time still an escape room? That is a question that a future escape room design study without the element of time might be able to answer.

Another possible follow-up study could be an escape room design research, which is tested in one or several university classes, led by a university teacher, with a cooling-down period designed to provide more confirmation and an elaborate discussion on the made decisions regarding the grey dilemmas. The test rounds regarding this design study got no teachers involved, but might prove interesting to investigate. The escape room is eventually going to be used to provide a research integrity orientation for master students. Then a university teacher also needs to know how the escape room works and how the activity and the cooling-down period should be moderated. Designing the cooling-down period as a focus group showed rather negative effects on the intrinsic motivation, especially regarding the participants' perceived competence. When designing the

cooling-down period more focused on the participants' interest, their interest and perceived competence towards the escape room discussion might increase.

11 Re-evaluation and design recommendations

Based on the research results, the design contained promising and elements that are in need of improvement.

The *promising* elements of this design:

1. *Time pressure*: besides the fact that time pressure drives students to put more effort in the activity, the participants linked the time pressure inside the activity to their own research practices and deadlines. However, the element also proved to have a negative effect on the grey dilemma discussions, decision making and research integrity choices in the rooms. Several students felt rushed and this caused to not all well-thought made decisions regarding grey dilemmas and integrity issues.
2. *Dialogues with grey dilemmas*: The dialogues and dilemmas were meant as a connection with the actual escape rooms. The current design did not get the best out of this element, because the dialogue contents were too different from the room contents. However the dialogues and dilemmas did generate quality discussions about research integrity. Whenever the connection between the dialogues and the rooms will be more clear, this could potentially increase the research integrity awareness inside the rooms as well. Another improvement point towards the grey dialogue dilemmas is giving the participants the option to generate own answers, rather than given options to choose from. This will increase the participants' feeling of autonomy and will decrease the frustration of choosing an answer they might not agree on.
3. *The introduction room*: Before the participants started on the actual escape room activity, they got send a PowerPoint file containing a short escape room with an introduction to the research integrity theme of the activity. The introduction room created for this design however was not clear enough. The introduction was either not working and most participants needed more explanation from the moderators to actually understand what was expected from them during the main activity. Even though, the element of an introduction is promising on its own, besides the ineffective use in the current design. The introduction element grants exactly the amount of information that you want the participants to know before starting the main activity. The introduction room is also able to provide the participants with insights in how the rooms will work in the main activity to give them a proper preparation possibility.
4. *Storyline/Narrative/Theme*: The aim of the storyline/narrative was to make the escape room experience as recognizable as possible to their own research practices. Regarding the current design and results, the storyline and visuals served its purpose in creating this relation to the real-life research practices. Participants were able to relate the activity to own research practices like writing a thesis or participating in a project.
5. *Increasing puzzle difficulty per room*: Every room in the design had its own difficulty level, which increased per room as you progressed through the activity. The first room was the most easy, with 10 minutes for the participants to complete. Room one was solved by one out of three groups. The second room was of medium difficulty, with 15 minutes for the participants to complete it. The third room was of hard difficulty, with 15 minutes for the participants to complete it. This room has not been solved at all (although the last group came very close). The last room was designed as a room with a rather low difficulty. A

difficulty in between the difficulty of the first and second room (easy-medium). The reason for this was to stimulate the confidence of the participants, after possibly not completing the third room. Not completing a room in time is a negative experience (failure) and the activity was designed to end on a positive note: with a high chance of completing the last room.

The elements of this design that are in need of *improvement*:

1. *The online setting (PowerPoint and Microsoft Teams)*: the PowerPoint file of this activity caused a lot of technical issues and problems with MacBook devices. The PowerPoints of Windows devices worked fine most of the time. Older devices had problems with the rather large PowerPoint files they had to download before the participants were able to participate in the first place. The Microsoft Teams environment caused some participants to feel distant to their group members, which influenced their experience as a whole. This mainly was caused because the cameras of the other participants were not visible because of the PowerPoint full screen mode.
2. *The cooling-down period*: The cooling-down period might be the most important phase in the learning cycle of the participant. Current design could not fully focus on the cooling-down period because of time pressure (ironically). There was just too much to discuss because the design was too big, with the activity already taking more than an hour of the participants' time. When focusing on learning results, the cooling-down period needs to be addressed and every choice that has been made needs to be discussed. The cooling-down period needs to be designed towards answering the participants' interest, rather than the interest of the researchers or teachers. This could generate competence and more interest of the participants towards the research integrity contents.
3. *The size of the activity*: the escape room design was too big to conduct research efficiently. It caused the researchers to also have less time to conduct the test rounds, because the design phase took too much time. Instead of four rooms with four dialogues, two rooms with two dialogues would have been the better design choice regarding the aim of the study. This would have caused the design phase to be halved, generating more time for the test rounds. For a lesson activity at the University, the size of the activity should be all right (when introduced effectively). Most educational escape rooms take one hour to complete. Without the technical issues, this activity could be completed within the sixty minutes (excluding the introduction and the cooling-down period).
4. *The third room (doctor's office)*: The doctor's office was very hard to complete, because none of the three participating groups did. The room was simply too hard to solve in 15 minutes. One of the reasons could be the limited conditions to communicate with each other in Microsoft Teams instead of physical communication. But compared to the other rooms, the doctor's office just had too many steps to complete and some clues were too hard to understand for the participants.
5. *The two integrity routes inside the rooms*: Every room in the activity contained two possible routes to complete it. One route took significantly less time than the other route, but was less integer. The other route took more time to complete, but was the most integer option. The design wanted to see the participants thinking about the routes and making thoughtful decisions, but instead no clear decisions were made by all groups. Most participants even told the designers that they did not notice the difference in integrity at all. Although it is a promising escape room element, looking at the current design it had not the desired effect. The difference between the dialogues/dilemmas and the rooms was just too big, looking at the research integrity discussions and awareness. Inside the rooms, the participants just

were not as aware and occupied with the research integrity content as during the dilemma discussions in the dialogues.

12 Acknowledgements

I would like to acknowledge the cooperation with master SEC student Sarah Seghier, for her helpful insights, critical views and creativity on the design process of the escape room activity. I would also like to acknowledge professor and supervisor Roald Verhoeff, and professor and project coordinator Mariëtte van den Hoven. Both supported us during the design process and helped greatly on approaching master students, which we needed for the three tests rounds. Lastly, I want to acknowledge the professors and members of the Escape Room Project for testing the escape room activity off record, to provide us with more insights and inspiration for the final design.

13 References

- Al-Azawi, R., Al-Faliti, F., & Al-Blushi, M. (2016). Educational gamification vs. game based learning: Comparative study. *International Journal of Innovation, Management and Technology*, 7(4), 131.
- Anderson, M. S., Shaw, M. A., Steneck, N. H., Konkle, E., & Kamata, T. (2013). Research integrity and misconduct in the academic profession. Paulsen & Perna, *In Higher education: Handbook of theory and research* (pp. 217-261). Springer, Dordrecht.
- Andorno, R. (2020, January 07) How is research integrity being taught in Europe? Retrieved from <http://h2020integrity.eu/how-is-research-integrity-being-taught-in-europe/>
- Arciniega, L. M., & Menon, S. T. (2013). The power of goal internalization: studying psychological empowerment in a Venezuelan plant. *The International Journal of Human Resource Management*, 24(15), 2948-2967.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191.
- Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. *Englewood Cliffs, NJ*, 1986(23-28).
- Bandura, A. (1997). Self-Efficacy: The exercise of control. New York, NY: *W. H. Freeman*.
- Borrego, C., Fernández, C., Blanes, I., & Robles, S. (2017). Room escape at class: Escape games activities to facilitate the motivation and learning in computer science. *JOTSE*, 7(2), 162-171.
- Buckley, P., & Doyle, E. (2016). Gamification and student motivation. *Interactive learning environments*, 24(6), 1162-1175.
- Carter, N., Bryant-Lukosius, D., DiCensio, A., Blythe J., & Neville, A. J. (2014). The use of triangulation in qualitative research. In *Oncology nursing forum* (Vol. 41, No. 5, p. 545-547).
- Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological bulletin*, 140(4), 980.
- Clarke, S., Arnab, S., Keegan, H., Morini, L., Wood, O. (2016). EscapED: Adapting Live-action, interactive games to Support Higher Education Teaching and Learning Practices, Games and Learning Alliance 2016: 44-153.
- Center for Self-Determination Theory (2021). *Intrinsic Motivation Inventory (IMI)*. Consulted on the 27th of March, from [https://selfdeterminationtheory.org/intrinsic-motivation-inventory/#:~:text=The%20Intrinsic%20Motivation%20Inventory%20\(IMI,target%20activity%20in%20laboratory%20experiments](https://selfdeterminationtheory.org/intrinsic-motivation-inventory/#:~:text=The%20Intrinsic%20Motivation%20Inventory%20(IMI,target%20activity%20in%20laboratory%20experiments)
- Costello, B., & Edmonds, E. (2007). A study in play, pleasure and interaction design. In *Proceedings of the 2007 conference on Designing pleasurable products and interfaces* (pp. 76-91). Association for Computing Machinery (ACM).

- Csikszentmihályi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper and Row, as cited in: Csikszentmihalyi, M., & Larson, R. (2014). *Flow and the foundations of positive psychology* (pp. 209-226). Dordrecht: Springer.
- Csikszentmihályi, M., & Larson, R. (2014). *Flow and the foundations of positive psychology* (pp. 209-226). Dordrecht: Springer.
- Cummins, J. (1986). Empowering minority students: A framework for intervention. *Harvard Educational Review*, 56, 18-36.
- Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of research in personality*, 19(2), 109-134.
- Deci, E. L., & Ryan, R. M. (2008). Self-determination theory and the role of basic psychological needs in personality and the organization of behavior.
- Deci, E. L., & Ryan, R. M. (2010). Self-determination. *The Corsini encyclopedia of psychology*, 1-2.
- Deci, E. L., & Ryan, R. M. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments, Tampere, Finland*, (p. 9-15).
- Dworkin, S. L. (2012). Sample size policy for qualitative studies using in-depth interviews.
- Eukel, H., & Morrell, B. (2020). Ensuring Educational Escape-Room Success: The Process of Designing, Piloting, Evaluating, Redesigning, and Re-Evaluating Educational Escape Rooms. *Simulation & Gaming*, 1046878120953453.
- Fanelli, D. (2011). The black, the white and the grey areas: Towards an international and interdisciplinary definition of scientific misconduct. *Promoting research integrity in a global environment*, 79-90.
- Frey, N., & Fisher, D. (2010). Motivation requires a meaningful task. *English Journal*, 30-36.
- Glavas, A., & Stascik, A. (2017). Enhancing positive attitude towards mathematics through Introducing Escape Room games. *Mathematics education as a science and a profession*, 281-293.
- Gómez-Urquiza, J. L., Gómez-Salgado, J., Albendín-García, L., Correa-Rodríguez, M., González Jiménez, E., & Cañadas-De la Fuente, G. A. (2019). The impact on nursing students' opinions and motivation of using a "Nursing Escape Room" as a teaching game: A descriptive study. *Nurse education today*, 72, 73-76.
- Harpine, E. C. (2015). Is Intrinsic Motivation Better Than Extrinsic Motivation?. In *Group-Centered Prevention in Mental Health* (pp. 87-107). Springer, Cham.
- Hoskonen, K. (2007). Mathematics is—favourite subject, boring or compulsory. In *Proceedings of the fifth congress of the European Society for Research in Mathematics Education* (p. 248).
- Huang, W. H. Y., & Soman, D. (2013). Gamification of education. *Research Report Series: Behavioural Economics in Action, Rotman School of Management, University of Toronto*.
- Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons.
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and social psychology review*, 2(3), 196-217.
- Kim, O., & Walker, M. (1984). The free rider problem: Experimental evidence. *Public choice*, 43(1), 3-24.
- Kinio, A. E., Dufresne, L., Brandys, T., & Jetty, P. (2019). Break out of the classroom: the use of escape rooms as an alternative teaching strategy in surgical education. *Journal of surgical education*, 76(1), 134-139.

- Kislenko, K., Grevholm, B., & Lepik, M. (2007). Mathematics is important but boring: Students' beliefs and attitudes towards mathematics. In *Nordic Conference on Mathematics Education: 02/09/2005-06/09/2005* (pp. 349-360). Tapir Academic Press.
- Lindahl, J. F., & Grace, D. (2018). Students' and supervisors' knowledge and attitudes regarding plagiarism and referencing. *Research integrity and peer review*, 3(1), 10.
- Ling, L. T. Y. (2018). Meaningful Gamification and Students' Motivation: A Strategy for Scaffolding Reading Material. *Online Learning*, 22(2), 141-155.
- McQuillan, P. J. (2005). Possibilities and pitfalls: A comparative analysis of student empowerment. *American Educational Research Journal*, 42(4), 639-670.
- Menon, S.T. (2001), 'Employee Empowerment: An Integrative Psychological Approach,' *Applied Psychology: An International Review*, 50, 153– 180.
- Michael, D., & Chen, S. (2006). *Serious games. Games that educate, train, and inform*. Boston, MA: Thomson.
- Moreira, P. A., Dias, P., Vaz, F. M., & Vaz, J. M. (2013). Predictors of academic performance and school engagement—Integrating persistence, motivation and study skills perspectives using person-centered and variable-centered approaches. *Learning and Individual Differences*, 24, 117-125.
- Murphy, K. R., & Aguinis, H. (2019). HARKing: How badly can cherry-picking and question trolling produce bias in published results?. *Journal of Business and Psychology*, 34(1), 1-17.
- Nacke, L., Drachen, A., Kuikkaniemi, K., Niesenhaus, J., Korhonen, H. J., Hoogen, W. M., Poels, K., IJsselsteijn, W. A., & De Kort, Y. A. (2009). Playability and player experience research. In *Proceedings of DiGRA 2009: Breaking New Ground: Innovation in Games, Play, Practice and Theory*. DiGRA.
- Nichols-Casebolt, A. (2012). *Research integrity and responsible conduct of research*. OUP USA.
- Nicholson, S. (2015). Peeking behind the locked door: A survey of escape room facilities. *White Paper available at <http://scott.nicholson.com/pubs/erfacwhite.pdf>*.
- Otis, N., Grouzet, F. M., & Pelletier, L. G. (2005). Latent motivational change in an academic setting: a 3-year longitudinal study. *Journal of educational psychology*, 97(2), 170.
- Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & education*, 52(1), 1-12.
- Pavlas, D., Heyne, K., Bedwell, W., Lazzara, E., & Salas, E. (2010, September). Game-based learning: The impact of flow state and videogame self-efficacy. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 54, No. 28, pp. 2398-2402). Sage CA: Los Angeles, CA: SAGE Publications.
- Polit, D.F., & Beck, C.T. (2012). *Nursing research: Generating and assessing evidence for nursing practice*. Philadelphia, PA: Lippincott Williams and Wilkins
- Rappaport, J. (1981). In praise of paradox: A social policy of empowerment over prevention. *American Journal of Community Psychology*, 9, 1-25.
- Rappaport, J. (1984). Studies in empowerment: Introduction to the issue. *Prevention in Human Services*, 3, 1-7.
- Robishaw, J. D., DeMets, D. L., Wood, S. K., Boisselle, P. M., & Hennekens, C. H. (2020). Establishing and Maintaining Research Integrity at Academic Institutions: Challenges and Opportunities. *The American journal of medicine*, 133(3), e87-e90.
- Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology*, 43(3), 450.
- Salen, K., Tekinbaş, K. S., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. MIT

press.

- Sawyer, B., & Smith, P. (2008, February). Serious games taxonomy. In *Slides from the Serious Games Summit at the Game Developers Conference* (Vol. 5).
- Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and engineering ethics*, 12(1), 53-74.
- Steneck, N. H. (2007). Introduction to the responsible conduct of research. *Washington, DC: US Government Printing Office*.
- Su, C. H., & Cheng, C. H. (2015). A mobile gamification learning system for improving the learning motivation and achievements. *Journal of Computer Assisted Learning*, 31(3), 268-286.
- Tranquillo, J., & Stecker, M. (2016). Using intrinsic and extrinsic motivation in continuing professional education. *Surgical neurology international*, 7(Suppl 7), S197.
- Turens, J. F. (2005). Teaching research integrity and bioethics to science undergraduates. *Cell Biology Education*, 4(4), 330-334.
- Van Bree, J., & de Lat, S. (2011). Complex systems and emergent behavior: Engaging with computer games to enrich organization studies.
- Van Bree, J. (2013). *Game Based Organization Design: New tools for complex organizational systems*. Springer.
- Van den Hoven, M., & Krom, A. (2020). Empowerment and Conceptual Clarity in Research Integrity: Comment to David Shaw, The Quest for Clarity in Research Integrity: A Conceptual Schema, *Sci Eng Ethics* (2019) 25: 1085–1093. *Science and engineering ethics*, 26(3).
- Wendel, V., Gutjahr, M., Göbel, S., & Steinmetz, R. (2013). Designing collaborative multiplayer serious games. *Education and Information Technologies*, 18(2), 287-308.
- Wiemker, M., Elumir, E., & Clare, A. (2015). Escape Room Games. *Game Based Learning*, 55.

14 Appendices

14.1 Appendix A: Observation form

Motivation category	Abbreviation	Description
Interest/Enjoyment		
Enthusiasm/Enjoyment	EN	The player(s) showed signs of enthusiastic behavior like laughing and smiling.
Interest in integrity issues.	IN	The player(s) showed interest in integrity issues. For example: long discussions about integrity dilemmas.
Perceived competence		
Acting with confidence when encountering integrity issues.	C1	When the participants encounter integrity dilemmas, they know what they want to do and what they think is the 'right' thing to do in the given situation and are communicating this with a confident attitude.
Recognizing integrity issues.	C2	The participants recognized integrity issues in the dialogues and in the rooms.
Solving integrity issues.	C3	The participants used their own research knowledge to solve integrity dilemmas and issues.
Puzzle solving skills	C4	The participants show integrity knowledge and skills while solving puzzles. For example, a participant shares guidelines from the code of conduct regarding referencing or citing.
Argumentation	C5	The participant underpins his/her opinion with (solid) arguments.
Effort/Importance		
Put effort in completing the activity.	E1	The participants work hard to complete the activity as good and as fast as possible. For example: the participants write information down on a piece of paper. Also participating to the discussions shows effort.
Persisted when stuck.	E2	The participant did not give up when encountering technical issues or problems.
Uninterested attitude.	I	The participants show uninterested behavior during discussions or while solving puzzles. They say things like: 'it does not matter what we answer'.
Pressure/Tension		
Showed signs of pressure/tension.	P	The participants showed behavioral signs of pressure and tension during the activity. For example: mentioning the timer frequently.
Perceived choice		
Option not available.	PC	The participants came up with own options for answering the dilemmas, because they would not use any of the options if they would have been in that certain situation themselves.
Relatedness		
Communication.	R	The participants communicate with each other. They respect and value their teammates opinions.
Value/Usefulness		
NVT.	NVT.	Not observable during the activity.

14.2 Appendix B: Observation example

Timespan (minutes)	Names	Codes	Categories	Quotes
00:16,00 – 00:17,00	Student 1 Student 2 Student 3 Student 4	EN, E1 (writing) E1 (writing) EN	Enjoyment Effort	
00:17,00 – 00:18,00	Student 1 Student 2 Student 3 Student 4	R, P R, P	Pressure Relatedness	‘Oh no!’ ‘We got x minutes left.’
00:18,00 – 00:19,00	Student 1 Student 2 Student 3 Student 4			
00:19,00 – 00:20,00	Student 1 Student 2 Student 3 Student 4	C4, PC C4, PC	Perceived competence Perceived choice	‘The choice we want to make is not an option.’
00:20,00 – 00:21,00	Student 1 Student 2 Student 3 Student 4	E2	Effort	
00:21,00 – 00:22,00	Student 1 Student 2 Student 3 Student 4	Quietly reading Quietly reading Quietly reading Quietly reading		
00:22,00 – 00:23,00	Student 1 Student 2 Student 3 Student 4	Quietly reading Quietly reading Quietly reading Quietly reading		
00:23,00 – 00:24,00	Student 1 Student 2 Student 3 Student 4	E2 (stuck) Crashed EN E2 (stuck)	Enjoyment Effort	
00:24,00 – 00:25,00	Student 1 Student 2 Student 3 Student 4			‘I would have a discussion with the supervisor about the matter.’ ‘I’m fine with that.’ (sounds uninterested)

14.3 Appendix C: Post-activity survey

Interest/Enjoyment

1. I enjoyed doing the escape room activity very much.
2. I would describe the escape room activity as very interesting.
3. This escape room activity did not hold my attention at all.

Perceived Competence

4. I think I did pretty well at this escape room activity, compared to other students.
5. After working at this activity for a while, I felt pretty competent in answering and recognizing integrity dilemmas.
6. I am satisfied with my performance at this task.

Effort/Importance

7. I put a lot of effort into this escape room activity.
8. It was important to me to do well at the escape room activity.

Pressure/Tension

9. I felt very nervous while doing this escape room activity.
10. I felt very tense while doing this escape room activity.
11. I felt pressured while doing this escape room activity.

Perceived choice

12. I believe I had some choice about doing this escape room activity.
13. I did this escape room activity because I wanted to.

Relatedness

14. I felt like I could really trust my teammates
15. I felt really distant to my team.

Value/Usefulness

16. I think this is an important activity.
17. I believe this activity could be of some value to me.
18. I think that doing this escape room activity is useful for _____

A Likert scale 1-7 was used on all questions, except 18 (open question):

Not at all true (1)			Somewhat true (4)			Very true (7)
---------------------	--	--	-------------------	--	--	---------------

14.4 Appendix D: Focus group instrument

A. *Discussing escape room (5 – 10 min)*

1. Clear dilemmas: all answers where possible, some of them more integer.
2. Hidden dilemmas: have you seen them all? Taken into account?
3. Rooms: intention per room?

B. *Questions on motivation and empowerment (30 - 40 min)*

1. How did you experience the escape room activity?
2. Was it a realistic experience for you to fit the shoes of the main character James?
3. Did you think the storyline was recognizable for you?
4. How did the collaboration go in the different rooms?
5. Was there a difference in cooperation in the rooms for you?
6. How did the discussions go in your opinion?
7. Do you feel that by participating in this activity, you could recognize research integrity dilemmas more quickly in given situations?
8. And in your own research?
9. And could you say that you could make a well-considered choice when you encounter such a dilemma?
10. How did you make the 'right' choice? (puzzle and dilemma separately).
11. What influence(s) did the following factors have on making the 'right' choice: pressure, drive, knowledge/skills etc.
12. Have you gained insights into what 'good' choices are in research integrity?
13. Do you feel like you have learned more about how to deal with research integrity dilemmas through this experience?
14. Can you explain why with an example?

C. *In conclusion (5-10 min)*

1. Describe in one sentence what your take-home message is from this escape room experience.
2. What is the most important thing you have learned from this experience?
3. What appealed to you the most in the escape room?
4. And what appealed the least?
5. Do you think that the escape room would be a suitable educational tool to bring research integrity to the attention of master students?

14.5 Appendix E: Narrative/Storyline practical elaboration

As stated in section 5.1.2 and 5.1.3, the escape room consists of four dialogues paired with a grey dilemma (with a twist) and four (escape) rooms filled with puzzles, (grey) dilemmas and tasks to solve by communicating and working together as a team. This section zooms in on the practical use of the dialogues and rooms.

The first dialogue describes a situation where the fictional students are in doubt about the hypothesis they drafted. Justin (fictional student) then comes with an idea: just wait with the construction of a hypothesis until after conducting the research. The participants then have to choose between four options if they think that Justin's idea is good, bad or maybe they do not know if it is a good or bad decision and they want to discuss it with their supervisor. When the participants made a decision, the dialogue continues with a twist. A form of external pressure appears in the form of for example the supervisor. This extra pressure could cause the participants to change their answers, but that would not in every case be the most integer thing to do. The first room follows after the first dialogue about the hypothesis, which is the student room. The fictional students made a choice about the hypothesis. In the room they will revise their reference list, before they will continue in the thesis process. In the room they need to reference via the 'right' source.

When the participants completed the room, the second dialogue begins where the free-riding dilemma is displayed. Justin just did not collaborated as much as James and Amber would have liked. Should James and Amber leave Justin out of the paper or should he get another chance? The twist shows that Justin apologizes for his actions and gives an explanation. His dad came home after being deployed by the army, which is why he could not contribute as much as he would have liked to. After deciding for the second time what to do with Justin, the second room starts. Justin has left the group (it does not matter what the group decides), James and Amber are the only ones left to complete the thesis. Inside the room, participants have to search for the feedback they got from their supervisor. They also have to use it according to the code of conduct rules, by solving puzzles and using their research knowledge.

Unlocking the door grants the participants access to the third dilemma. The feedback is used to improve the thesis and now it is time to conduct the research and collect data. The third dialogue describes a choice between conducting own research about BMI or using data out of an already existing data-set of 20 patients, generated one year ago. The difference is that the research the participants have to conduct, only consists of 12 patients and is rather time consuming in comparison to the data-set. The only problem with the data-set is that it has an incomplete logbook. The participants have to choose whether they will take the data-set or if they conduct their own research or something in between, like consulting the supervisor about the matter. The twist after the decision then reminds James and Amber through the supervisor about the limited time they still have, because third parties are eager to see results. The participants move on to the third room, the doctor's office, after providing an answer to the same dilemma after seeing the twist. Inside the room they have to complete their own data (even if they choose differently in the dialogues) by solving puzzles and completing tasks together, or they could take the easy way out by using the data-set that is mentioned in the dialogue.

Entering the final code in the scale grants the participants access to the fourth and final dialogue. This dialogue is about the data that has been collected, but this data is flawed. The scale was not calibrated and there is no time left to redo the research before the deadline. What to do? When deciding which of the four options suited the most, the twist appears: the director of the hospital where the participants conducted their research. He is aware of the fact that the data is flawed and (firmly) requests that the data is not going to be used for publication, because that could

damage the reputation of the hospital. After deciding what to do with this twist, the participants progress to the last room on data analysis. Their goal is to conduct the statistical analysis on the data in the form of a SUM-score. They need to find and add together three variables in the room, which will give the answer to the SUM-score and therefore the final code of the room. The participants are also able to choose the easy way out, by picking up the answer from a discussion board on Blackboard, which is less integer because the statistical analysis cannot be checked. The final code prints the thesis on the printer and completes the escape room activity.