Universiteit Utrecht

Investigating Research Master students' interests in Research Integrity to improve its education

By: Romy Zandstra (5619505)

Supervisors: Alice Veldkamp and Roald Verhoeff

16-06-2022

Utrecht University

Abstract

Research Integrity (RI) covers ethical concerns during the different stages of the research cycle. RI education has gained attention in response to a series of scandals of misconduct, and means to teach students Responsible Conduct of Research (RCR). However, current RI education is not rated highly and considered boring by Research Master (RMA) students at Utrecht University, jeopardising its ability to reach its goals.

This research aims to identify topics within RI that RMA students find most interesting, so RI education can then be shifted to focus more on those topics, hopefully reawakening RMA students' interest in RI. A secondary aim was to discover if a RMA student's Graduate School was an important determinant for their interest in certain topics within RI.

The study consisted of two parts, a questionnaire to collect quantitative data on which topics within RI are important to RMA students of Utrecht University, and then semi-structured interviews to gather qualitative data on why these topics are considered relevant or interesting. The questionnaire received 114 responses from students of the Graduate Schools of Humanities, Natural Sciences, Life Sciences, Teaching, Social and Behavioural Sciences, Geosciences and Pharmacy. The interviews numbered four in total, with students from the Graduate School of Teaching.

The questionnaire revealed that topics on communicating with and support from supervisors, ethical handling of both quantitative and qualitative data (from safekeeping participants' privacy to selecting proper statistical methods), collaboration with peers and referencing correctly are perceived as most important.

Interviews and qualitative data from the questionnaire showed that especially the first two topics (supervisor and data) are considered important, due to the large impact both have on the eventual success and graduation of a student. The 'collaboration with peers' topic received more diverse responses, with some students finding it unnecessary to further learn about, while others do find it important. 'Referencing' was generally considered an important topic, but not *as* important as the other three.

When looking at differences between the different Graduate Schools, the topic 'data' seemed most divisive. There were both major differences in what kinds of data (qualitative or quantitative) students from different Graduate Schools were familiar with, resulting in different ethical concerns being raised, but there was a differing level of guidance required as well. Maths-related RMA students do not collect or have to work with any type of data for their research, whereas the Graduate School of Geoscience students are taught hardly anything regarding data, as well as RI in general and referencing, so they have great need for education regarding these topics.

Keeping these nuances in mind, the results imply that RI educational developers should focus on flexible education and include several different ethical dilemmas regarding the four topics of supervisor, data, collaboration with peers and references, with the RI teacher then selecting the appropriate ones for the students and maybe sub-dividing the students into different groups based on interest.

Table of Contents

Introduction

Back in the late 1970s and 1980s, scandal after scandal of misconduct in the research community reached the headlines of newspapers. Several of these misconducts were absolutely massive in their scope, e.g. John Darsee with over seventeen falsely published papers and 50 abstracts (Relman, 1983), or Elias Alsabti who published 60 plagiarised papers (Broad, 1980). Several cases of alleged misconduct also had massive public impact, with testimony before US Congress and published books on both Thereza Imanishi-Kari and Robert Gallo (Kalichman, 2013).

Eventually, National Science organisations determined that these scandals were hurting the image of science in the public eye, and something had to be done to prevent research misconduct (National Academics of Science, 1992; Steneck, 2006). There was an assumption at the time that students would naturally learn how to be a good researcher from a mentor, but that turned out not to be the case (Swazey & Anderson, 1998) due to a variety of possible reasons (Kalichman, 2013). So, formal courses were created to teach students how to be a good researcher (Kalichman, 2013), a field now known as Research Integrity (RI). So, what is RI exactly, and what does RI education look like?

RI in literature is sub-divided into three categories: Responsible Conduct of Research (RCR), Falsification, Fabrication and Plagiarism (FFP), and Questionable Research Practices (QPR) (Steneck, 2006). These three categories will be explained further in the Theoretical Background, but here is a quick overview. RCR is the ideal way researchers *should* behave, any sanctionable actions are gathered under FFP, and due to reality being rather messy, there is a large grey area in between RCR and FFP called QPR (Steneck, 2006).

RI education is a very complicated topic, because while there are national standards, how to attain these standards is up to the teachers as there usually is no formal curriculum in a country on this subject (Kalichman, 2013). This means that many researchers found that teachers had different objectives and goals when it came to RI education (Kalichman, 2013; Kalichman & Plemmons, 2007; Antes & DuBois, 2014; Shepard et al., 2015). These different teaching practices then made it next to impossible to research whether RI education is actually effective on large scale (Kalichman, 2013; Resnik, 2014; Antes & Dubois, 2014). Despite RI education existing, misconduct is happening still, and in the month of March 2022 alone, four cases of misconduct were reported to the US Office of Research Integrity (ORI - The Office of Research Integrity, 2022). Clearly, RI education is still necessary.

Theoretical background

What is RI?

RI is at its essence any and all ethical concerns that are part of the research cycle, from design to publishing. As mentioned in the introduction, RI is generally divided into three categories in RI literature; Responsible Conduct of Research (RCR), Falsification, Fabrication and Plagiarism (FFP), and Questionable Research Practices (QPR) (Steneck, 2006). RCR is considered the ideal and responsible behaviour researchers should exhibit and is based on four main principles:

- "**Reliability** in ensuring the quality of research, reflected in the design, the methodology, the analysis and the use of resources.
- **Honesty** in developing, undertaking, reviewing, reporting and communicating research in a transparent, fair, full and unbiased way.
- **Respect** for colleagues, research participants, society, ecosystems, cultural heritage and the environment.
- **Accountability** for the research from idea to publication, for its management and organisation, for training, supervision and mentoring, and for its wider impacts." (ALLEA, 2017, p.4).

On the complete opposite end, FFP consists only of the three actions contained in the name, so Falsification, Fabrication and Plagiarism, and is considered completely undesirable. It is also heavily sanctioned, due to the damage these actions can do to both to research itself, and its image to the general public (National Academies of Science, 1992; Steneck, 2006).

In the middle of these two extremes lies QPR, which is a big grey area. Actions and behaviours belonging to this category do not comply with the four principles of RCR, but the National Academies of Science (1992) determined that they were not egregious and damaging enough to sanction. The vast majority of day-to-day ethical situations researchers face fall into this category, since these situations are often complex and cannot be fully answered by avoiding FFP and following the RCR guidelines. Think of arguments about authorship of a paper, lack of contact with a supervisor for students, or separating results from the same research into several papers just to have more published papers to your name.

RI education

As stated in the introduction, the foremost goal of RI education is to teach (future) researchers Responsible Conduct of Research, in order to minimise cases of misconduct (FFP). In practice and in different countries or institutions however, it can take different forms, with many different aims and goals.

Antes and DuBois tried to streamline these goals in 2014, so that research on the effectiveness of RI education could actually be conducted, and so teachers would have a clearer idea on how to teach RI. They expanded on the foremost goal of RI education, and made the point that RI education should in the end lead to better ethical decision-making, i.e. researchers and students are able to make better and more informed decisions on ethical matters, leading to RCR.

For a practical example of RI education which keeps this goal in mind, one can look towards the Massive Open Online Course (MOOC) on RI developed by teachers and experts on RI at Utrecht University for Elevate Academy. The writers of the MOOC want to approach RI from a positive standpoint and help students make the right decisions. There is therefore a lot of focus on both RCR and QPR, rather than FFP.

Key to this MOOC are ethical dilemmas, which are short descriptions of various situations in which an ethically correct decision needs to be made. Some of the ethical dilemmas provide several options for the student to choose from, while for others the student needs to think of a solution themselves. An example from the MOOC (adapted from the EUR Dilemma Game, nr. 18) is as follows:

No luck

I have run an experiment, but the results did not work out. I am disappointed because I had carefully designed all the manipulations and stimuli. The previous experiments that I ran for my project had worked out, and I had expected the outcome of the last experiment to be in line with the previous results. I am now writing my thesis, and I decide to write it on the basis of the successful experiments and to ignore the failed experiment.

This ethical dilemma clearly falls in QPR; the "I" figure in the ethical dilemma is not engaging in falsification or fabrication so it is not FFP, but is also not honest about their full results, meaning they are not behaving like a responsible researcher. Through discussing this ethical dilemma, students not only get familiar with situations they might come across during their career as a researcher, but also learn how to behave responsibly when confronted with situations like these.

Unfortunately, while current physical RI education at Utrecht University makes use of these very useful ethical dilemmas, it is not rated highly by students, specifically Research Master (RMA) students. For instance, in an RI course evaluation in 2019, Life Science students scored it on average three stars out of five, and many found it boring and uninteresting (van den Hoven, 2020). So, despite the importance of RI education, both in preventing misconduct but also in helping students improve their ethical decision making, RMA students are on average uninterested. How can this problem be solved?

Possible solutions

There are several possible approaches for this problem of students' negative perception of RI education. Teachers and researchers can for instance design a whole new educational tool to make learning about RI more fun for students, such as an educational Escape Room (Van den Hoven, 2020). Educational Escape Rooms have gained in popularity recent times, since they are originally meant for recreation and thus are a fun way for students to learn about a topic (Veldkamp et al., 2020).

Another approach, is figuring out which parts or **topics** within RI RMA students find interesting, to make the content more relevant to students. This thesis will focus on the latter solution.

RI covers any ethical concerns that can be part of the entirety research cycle, so the topics within RI are numerous and can often be further subdivided, e.g. the main topic of data and then subtopics like data management, participants' privacy and open source data. However, some may be considered more important by students than others, either because they have experienced an ethical dilemma regarding a topic, or because they realise the ramifications of a topic should something go wrong, or a mixture of both. To truly grab students' interest and make a course on RI interesting, it might then be best to make a selection in the teaching material to discuss only or mostly those topics students find important and/or have had to deal with (often). Since ethical dilemmas are an important part of teaching material for RI education and need to be selected for a course anyways, they might be a good target to make this more precise selection for.

However, information on which RI-related topics RMA students find important scarce. Most research on RI and students is done on PhD students, not RMA students (e.g. Jensen et al., 2018; Hofmann & Holm, 2019). Research that *does* concern RMA students (or Graduate students) often investigates other things than which topics within RI students find interesting, such as which factors influence for RMA students' ethical decision making (Langlais & Bent, 2014).

This leads me to the following Research Question:

Which kind of ethical dilemmas should be selected for Research Master students to make Research Integrity education at Utrecht University more interesting?

To determine what kind of ethical dilemmas should be selected, it is important to know which topics within RI RMA students find important. I also had an interesting experience within my studies, since I switched from a beta sciences Bachelor to a Social Sciences Master. This was a major shift and I suddenly experienced (amongst other things) different ethical issues than before. I wondered therefore if a student's Graduate School had any impact on the overall student population interest in RI. This leads me to the following Research sub-Questions:

1. Which Research Integrity related topics do Research Master students find relevant?
2. What differences can be found in the articulation of Research Integrity related topics amongst the Graduate Schools of Humanities, Natural Sciences, Life Sciences, Teaching, Social and Behavioural Sciences, Geosciences and Pharmacy?

Materials and Methods

The research consisted of two parts: a questionnaire and an interview. To provide an overview of which RI-related topics RMA students at Utrecht University find important, we created and administered a questionnaire. To determine *why* certain RI-related topics are perceived as relevant, we subsequently conducted semi-structed individual interviews. This way, we could both gauge the student's personal *interest* in an RI-related topic, as well as grasp their personal *experience* with the topic, since both seemed to be determinants according to the questionnaire results.

Research Master students studying at Utrecht University were the participants selected for both instruments. Questionnaire participants received an email from their RMA coordinator asking them to participate, and I also sent an email to my student association B.I.T.O.N. asking RMA students to participate. Interview participants on the other hand were recruited through personal messaging apps (Whatsapp) and were all people I already knew and was friendly with.

Questionnaire

The questionnaire consisted of four parts, described in detail below. It mostly collected quantitative data, but the final part collected qualitative data. For a full list of questions, see Appendix A. The questionnaire was designed by myself and two experts on RI, Mariëtte van den Hoven and Roald Verhoeff, based on their expertise and my personal experiences. The questionnaire was constructed in the following four parts:

1. Part one asks the students once more for consent and for their RMA

2. Part two poses seven ethical dilemmas regarding RI to students, which the students need to rate on a 5-point Likert scale, with one being highly relevant and five being not relevant at all. The purpose of this part of the survey is to refamiliarize the RMA students with RI, since RI education often comes in the form of ethical dilemmas, and to help prepare them for part three.

   The ethical dilemmas were specially formulated for this survey. Of specific concern were two things; the length and identification of a clear problem. The length of the ethical dilemmas was very important, the dilemma could not be too long or otherwise students might lose interest. A clear problem identification such as an ethical dilemma ending with "I am not sure what to do" would (hopefully) prompt the respondent to either find the dilemma familiar/recognisable or not. These initial formulations were also tested twice, explained further along this section.

   As an example of what this would look like practically is as follows, the others can be found in Appendix A:
   I am finishing a paper, and need a few references that confirm my statements. I choose a few articles that only briefly mention something about my paper topic, but I don't have the time to search for new references, so I decide to use them anyway.

3. Part three asks students three questions regarding sixteen different RI-related topics, with either a topic on its own or two or more subtopics. An example of the former would be "use of literature: incorrect or invalid referencing", an example of the latter "supervision: lack of communication". A full list of the topics can be found in Appendix A. These topics were once again constructed from my personal experience and the experience of the two experts on RI, with also advice from RMA coordinators on RI-related topics their students have to deal

with. These topics were also tested twice for viability, see below. The three questions about these topics are as follows:

- The frequency with which the students experienced these 16 RI-related topics, rated in a 5-point Likert scale from often to never.
- The personal Importance of these topics to the students in that moment, so as an RMA student, rated in a 5-point Likert scale from highly important to not important at all.
- Whether the students would like to learn more about these topics, answerable with yes or no.

These three questions were asked simultaneously per RI-related topic, so frequency – 5 answer options, next to it personal importance – 5 answer options and ending with willingness to learn more – 2 answer options. This meant that students could immediately answer all three per RI-related topic, instead of finishing one sub-question for sixteen RI-related topics, scrolling down and then having to do it twice more. From personal experience I know this is very frustrating and makes me want to quit questionnaires, so to avoid that I elected to ask the three sub-questions simultaneously.

The frequency of a topic can show which RI-related topics occur very often and thus might be important to discuss during an RI course. The personal importance a student thinks an RI-related topic holds, can also help identify suitable topics for an RI course. Finally, a student might experience an RI-related topic often and consider it personally important, but they might consider their current education regarding it sufficient, so they can still hold no desire to learn more about it. All in all, the three sub-questions reveal different interests and priorities a student can have towards an RI-related topic, and to get the full picture I decided to ask all three.

4. Part four consists of two open questions, the first asking students which of the 16-related topics they find most important and for them to explain why in their own words. The second asks the students whether they have experienced any other RI-related topics not mentioned in the rest of the survey. I asked the first question to identify the RI-related topics students find *especially* interesting and important, which are then the ones that should absolutely be included in a course on RI. The second question was meant to identify any topics students found important, but I missed.

Before the survey was widely distributed, it was tested by both RMA coordinators (three) and students (two). The RMA coordinators belonged to the Natural Sciences, Life Sciences and Literature Graduate Schools. They checked to see whether the balance in the survey was correct and whether the survey would be able to be answered by their students. They checked for things like how recognisable the ethical dilemmas were, whether they had an appropriate length, the selection of RI-related topics and whether anything was missing.

The students checked whether they would actually be able to answer the questions or found some items strange, and based on the time it took them to complete the survey, an estimated time was added to the recruitment email.

The survey was conducted online via Qualtrix, with a link to the survey and a consent form as part of the recruitment email. The total response was 114 RMA students, which was a satisfactory number.

Interview

I conducted interviews after the questionnaire results were analysed, to collect qualitative data on *why* certain RI-topics scored so high. I selected four RI-related topics for the interview (see Results or Appendix C for which ones).

I conducted four interviews in total, all with RMA students from the Graduate School of Teaching. Two of the participants were female, one was non-binary and one was genderqueer. I initially wanted to conduct interviews with a mix of students from different Graduate Schools and a mix of genders, but I contacted many and unfortunately students from different Graduate Schools were unavailable. Two of the interviews were conducted online, and two were in person, depending on the personal preference of the student. All interviews were recorded.

For each of the four topics, I asked the students both for their personal experience regarding that topic, and whether they found the topic important enough to learn about during a course on Research Integrity. For both of these main questions I went rather in depth, for the personal experience one I asked things like how often they had to deal with the topic, whether they had memorable memories regarding the topic and pinpointing why that memory was memorable. For the education section I then really went in depth on *why* they would or would not like to learn about a topic, and whether any sub-topics (for instance data – data analysis, data collection, data protocol, privacy of participants, etc.) were of particular (dis)interest.

The final question asked them to rank each of the four topics in order of importance and why this is the case. This would really condense their answers throughout the rest of the interview and give a good summary of the student's thoughts. See Appendix C for a full list of the questions.

Data analysis

The data analysis consisted of several different approaches. First of all, the questionnaire (mostly) consisted of 5-point Likert scale questions. Since there is debate in the scientific community on whether it is valid to conduct statistical tests on Likert scale questions, due to their short range, I mostly used simply statistics in the form of calculating percentages.

Due to the amount of data, I decided to subdivide it as follows per Research sub-Question: the ethical dilemma results, and the results of the three sub-questions regarding 16 RI-related topics were shown in four separate 2D 100% stacked bar charts, to give a good overview of the Likert scale answers together with percentages. The graph that showed the number of students belonging to each Graduate School and the one categorising the answers to the first open question were both cluster columns. These graphs were then interpreted manually. The total amount of graphs is then ten.

To answer the second Research sub-Question regarding differences between Graduate Schools however, I conducted a statistical test on my results, to condense my results and see any clear differences. So, since my data was ordinal and not normally distributed, I selected the Kruskal-Wallis H test. I decided to show all the rankings of the Graduate Schools for all the question items in a cluster column graph, not just any statistically significant results, due to the low number of participants of two Graduate Schools causing large outliers.

The interview data analysis was more informal, since I decided to combine the data of my interviews with the qualitative data from the questionnaire, both for more qualitative data in total, as well as a better overview of the entire student population's opinions on the most relevant RI-related topics. I chose to present this data in summaries of the general student population opinion along with representative and demonstrable quotes.

I therefore did not code my interviews, since the coding of the interviews would not match the answers given by the respondents of the questionnaire, making it difficult to analyse the sources together. I also did not make full transcripts of the interviews. Instead, I re-listened to the recordings of the interviews and created notes wherein I wrote along with the interviewee's answers and timestamped at which point in the recording each answer occurred. I did not write down my own interjections and may have occasionally summarised an answer, but I did ensure all quotes used in the Results section are the full interviewee's answers.

Results

The results for this study will be sub-divided per Research sub-Question. First the sub-Question regarding relevant RI topics will be answered, then the sub-Question regarding whether a RMA student's Graduate School is a determinant for their interest in RI-related topics. The results will be subdivided further into first the quantitative questionnaire data analysis and then the qualitative data (both interviews and the qualitative portion of the questionnaire) analysis.

Relevant RI-related topics for RMA students

Respondents

The total number of responses for the questionnaire was 114. How they were distributed amongst Graduate Schools will be discussed when answering the second sub-Question. The interviewees numbered four in total, two women, one non-binary and one gender queer person.

Quantitative analysis (questionnaire)

Reliability analysis

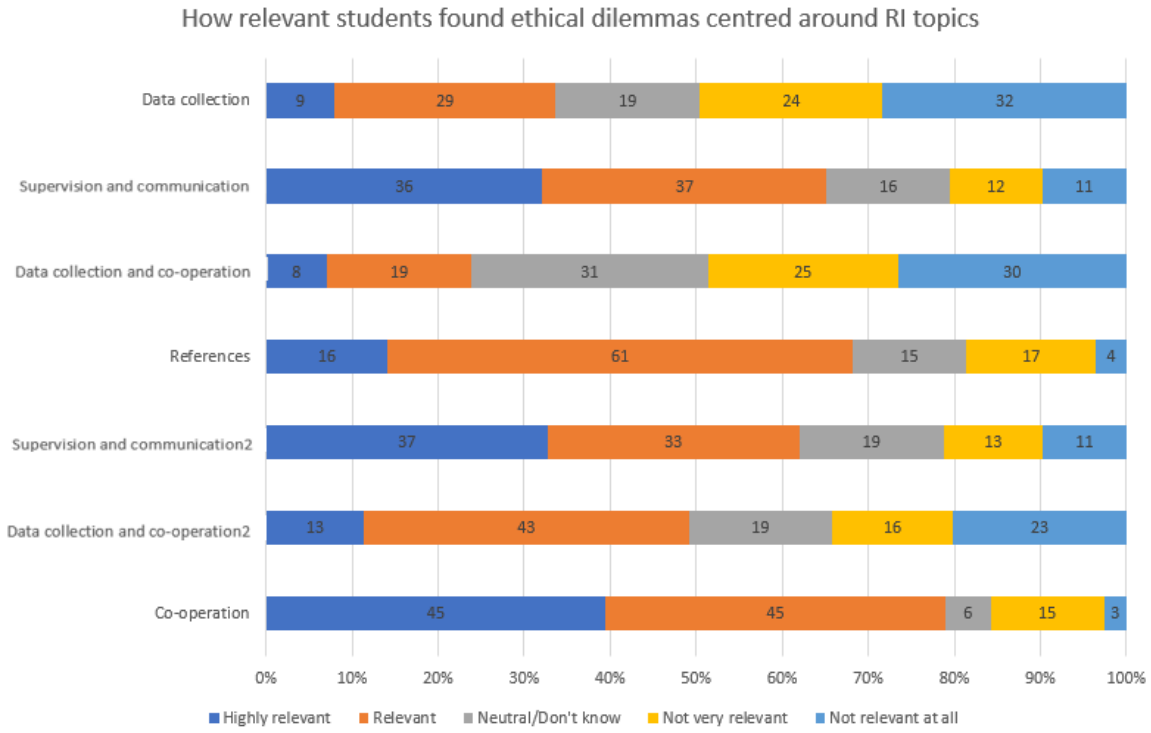
Cronbach's alpha can measure the reliability of a questionnaire through determining the correlation between items. The higher the number (from 0 to 1), the higher the reliability of the questionnaire and the more valid the results are. For this questionnaire, the Cronbach's alpha was 0.899 with N=55. This is very high, meaning the survey is very reliable.

Ethical dilemmas analysis

As stated in the Materials and Methods, the second part of the questionnaire consisted of short ethical dilemmas (see Appendix A for full descriptions). All these ethical dilemmas centred around specific aspects of RI, which can be seen in the graph below.

Figure 1

Percentage of students that perceived ethical dilemmas to be relevant



Note. Percentage of students that perceived ethical dilemmas to be relevant. The x-axis show the percentages of students who answered a certain Likert scale option, see legend. The y-axis shows the topic(s) of the seven ethical dilemmas.

Out of these seven ethical dilemmas, the ones concerning co-operation (78.9%) and both supervision and communication (65.2% for the first, 61.9% for the second) score highest on both highly relevant and relevant, while references (68.1%) also scores highly, but is mostly answered with relevant (54%). The lowest scoring topics are the first data collection and co-operation (23.9%) topic, and the data collection (33.6%) topic.

All the topics scoring highest on both highly relevant and relevant have to deal with co-operation and communication. Why this is the case and why references is mostly answered with relevant will be discussed later in the Results section and in the Discussion.

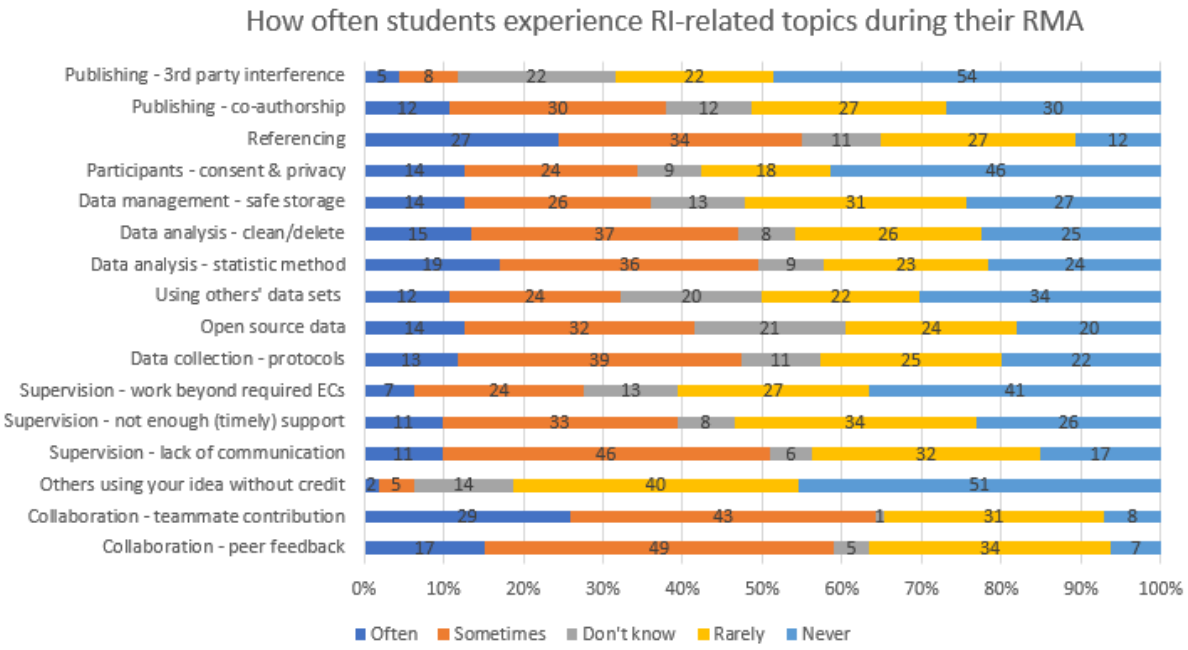
RI-related topics analysis

This question is a very large one, where I asked RM students to answer three different questions about sixteen RI-related topics. These three things were as follows: how often they came across the RI-related topic in their RM, how much the topic was personally important to them, and whether they would like to learn more about the topic. For easier analysis, these three sub-questions have been given their own graph and analysis.

Frequency

Figure 2

The frequency with which students experience 16 RI-related topics during their RMA



Note. The frequency with which students experience 16 RI-related topics during their RMA. The x-axis show the percentages of students who answered a certain Likert scale option, see legend. The y-axis shows the 16 RI-related topics.

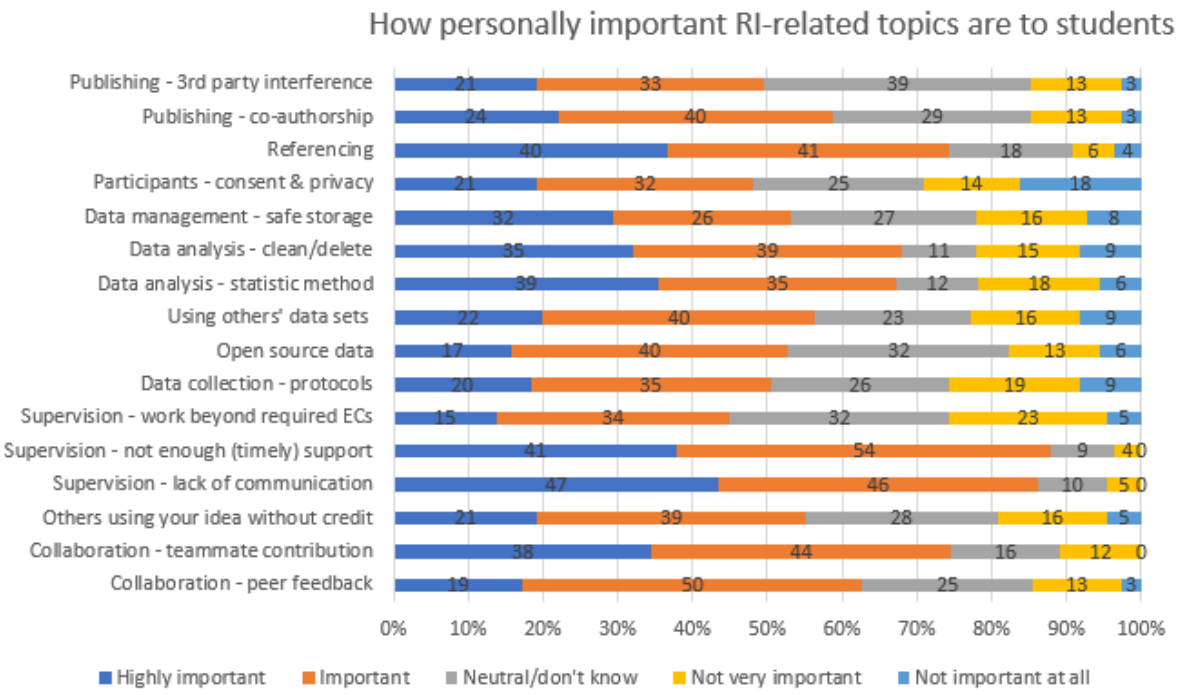
Out of these sixteen RI topics, collaboration peer feedback (58.9%), collaboration teammate contribution (64.3%), and referencing (55%) score highest when the often and sometimes responses are combined. The lowest scoring topics are others using your ideas without credit (6.3%) and 3rd party interference in publishing (11.7%).

The three most occurring topics are similar to the most relevant ones in the short scenarios, apart from the supervisor topics, which might be one of the reasons why these topics are considered so important.

Personal importance

Figure 3

How personally important the 16 RI-related topics are to students



Note. How personally important the 16 RI-related topics are to students. The x-axis show the percentages of students who answered a certain Likert scale option, see legend. The y-axis shows the 16 RI-related topics.

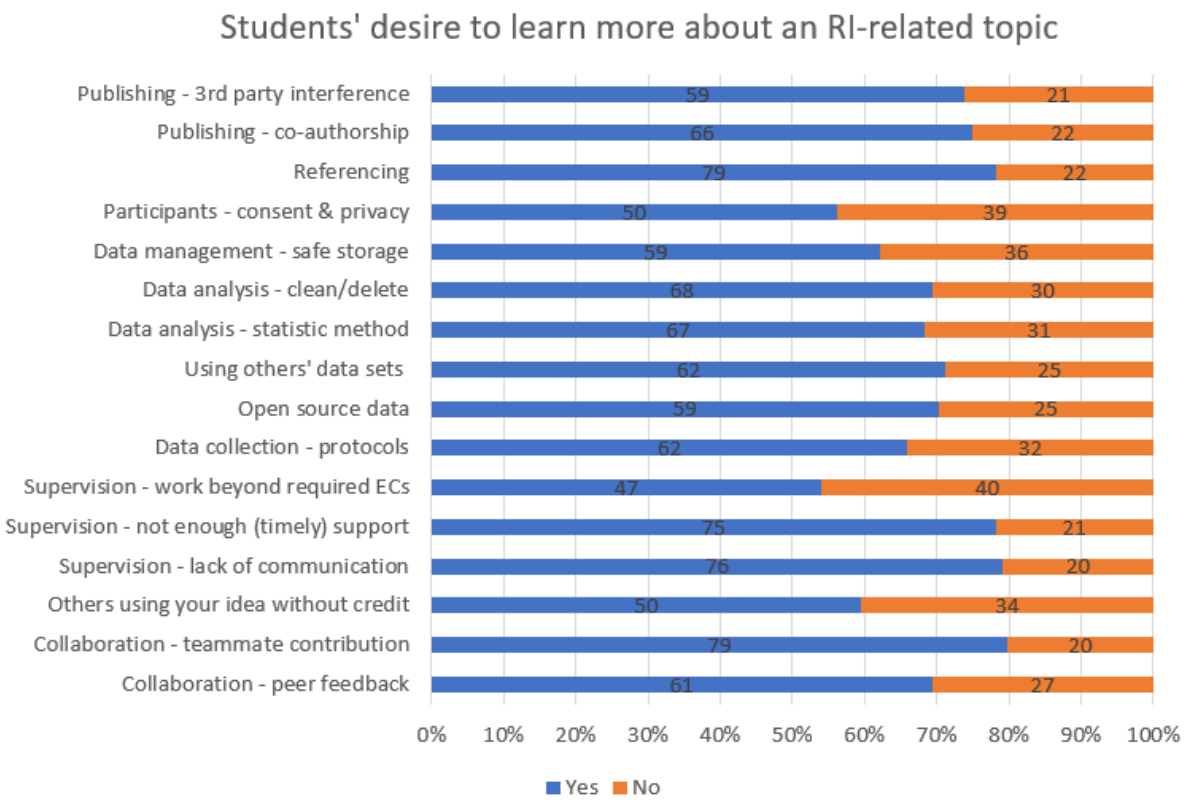
Out of these sixteen RI topics, collaboration teammate contribution (74.5%), supervision lack of communication (86.1%), supervision not enough (timely) support (88%), and referencing (74.3%) score highest when the highly important and important responses are combined. The lowest scoring topics are supervision work beyond required ECs (45%) and participants consent and privacy (48.1%).

There is a real trend emerging in which topics are considered most important and relevant, although all topics for this sub-question score rather high, with the lowest being 45%.

Desire to learn more about a topic

Figure 4

Students' desire to learn more about an RI-related topic



Note. Students' desire to learn more about an RI-related topic. The x-axis show the percentages of students who answered a certain Likert scale option, see legend. The y-axis shows the 16 RI-related topics.

Out of these sixteen RI topics, collaboration teammate contribution (77.8%), supervision lack of communication (79.2%), supervision not enough (timely) support (78.1%), and referencing (78.2%) score highest in students' desire to learn more about the topics. The lowest scoring topics are supervision work beyond required ECs (54%) and participants consent and privacy (56.2%).

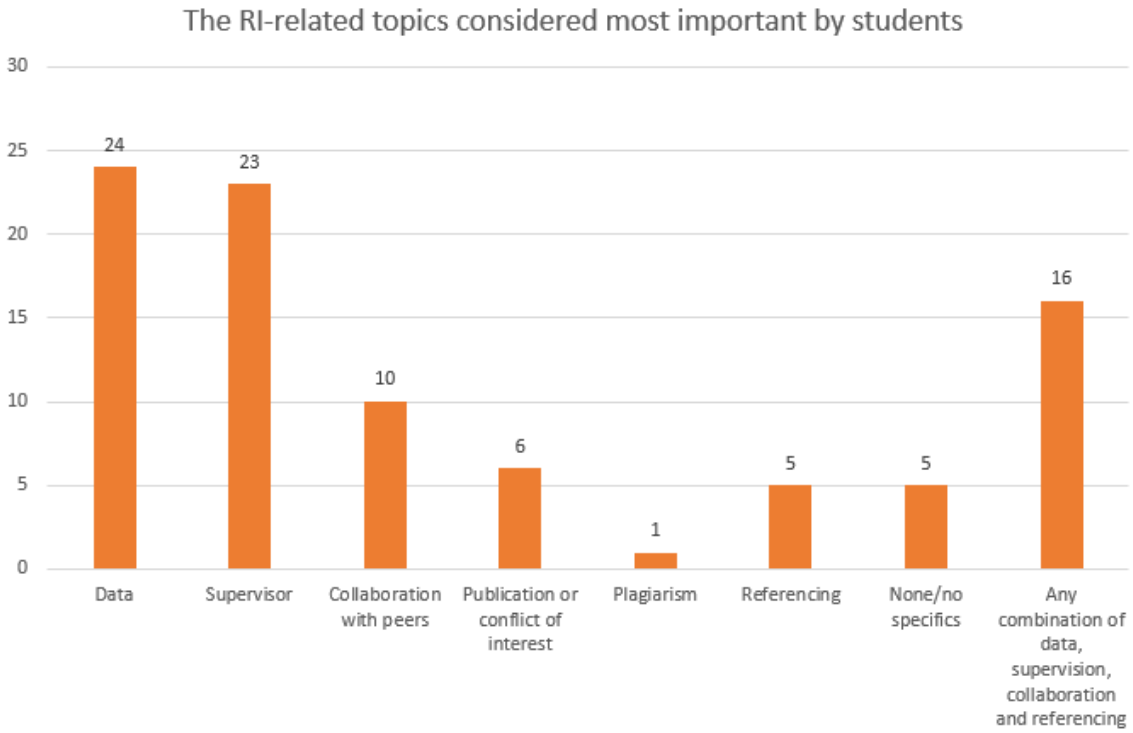
This sub-question's top four and bottom two topics are identical to the previous sub-question's, further increasing the trend seen among the other questions and strengthening my overall conclusions.

Open question quantification

To ensure I was not missing any interesting topics to include in the interviews, I decided to quantify the answers to the first open question of the questionnaire. This question asks the RMA students which topic out of the 16 RI-related topics in the previous question they found most important.

Figure 5

The 16 RI-related topics considered most important by students



Note. The 16 RI-related topics considered most important by students. The x-axis shows the possible the categories the answers to this open question could be assigned to. The y-axis shows the number of students who answered per category.

Out of eight categories I created, data (26.7%), supervisor (25.6%) and a combination (17.8%) are mentioned most often. The mentioned topics mentioned the least are plagiarism (1.1%) and referencing (5.6%).

This is the first time the data topic scores highly in the questionnaire, along with the usual topics of supervision, referencing and collaboration. However, since this question asked RMA students which RI-related topic they found *most* important, I decided to include all four in the interviews.

Qualitative analysis

I have two sources of qualitative data for my research; one is the interviews with four RMA students, and the other source is the two open questions of the survey filled in by 114 students. I decided to analyse both sources together, to get a good overview of the opinions of the entire student population for each topic. First I will discuss per topic why the RI-related topics of supervision, data, collaboration with peers and references are considered important to the interviewed RMA students, and then what RI education regarding these topics should and should not contain. Finally, I will discuss the opinions of the questionnaire respondents.

Supervisor

The RI-related topic of supervisor is considered very important by all four students and was ranked highly among the four interview topics, since “a supervisor is supposed to help students, and that help is often necessary for a student’s success.” One of the students also mentioned that: “Help is also an important part of succeeding, many students in this masters don’t know how to proceed in social science courses, it’s all really new, so if the supervisors have an experience with that, it is important to tell the student how to succeed.”

Most of the students have had good experiences with supervisors, but one had a supervisor for a course who “only told us what was wrong in his feedback, it wasn’t constructive at all.”, and found the experience to be very frustrating and demotivating.

When the students were asked whether the topic should be in a course on RI, all students agreed. They are generally interested in how to deal with a supervisor who doesn’t respond well or seem very invested in the student, how to choose a good supervisor, what the agreements are amongst the department for supervisors, and how to navigate the power (im)balance between students and their supervisors.

The importance of the supervisor topic is also reflected in the questionnaire, with students commenting things like: “I think the supervision-related issues are the most relevant, because as a student it is hard to criticize someone that decide [sic] on your grade. Even though I have had mostly positive experiences within this masters [sic] program, and have also heard positive things from others, I think it is the most relevant one because when you do have problems, it is hard to find out what to do.”

And: “No timely supervision. I once had to write a review and the supervisor agreed to give some feedback on my final version. However, he sent it to me 3 months after I handed it in and I missed my deadline because of it.”

Data

The data topic will briefly be discussed here, since there were large differences regarding Graduate Schools which will be discussed while answering the second Research sub-question. Data is considered very important by all students, and was consistently ranked either first or second (with supervisor then being first). The students’ biggest concerns regarding data were all centred on qualitative data, for instance regarding how to appropriately interact with research participants and privacy concerns regarding data storage. One student was however also interested in open source data, “since I haven’t heard about it yet, it is not a part of Research Methods or Research Proposal”.

This is very different from the students in the questionnaire, who were more focused on quantitative data and the issues that kind of data brings with it: “data analysis: unsure about statistical method; this is what I fear most for my upcoming master thesis” and “Data analysis & management. Will be very important both for thesis and the future” are two examples of students worrying about how to properly and ethically deal with their data.

Collaboration with peers

The opinions of the four students were very divided regarding this topic. Three students mentioned negative experiences where several group members did not properly contribute, but the same students then all agreed that collaboration is a skill you should already possess during the Master, and thus devoting a lot of time to it during a course on RI would not be very valuable. There is a particularly poignant quote from a student showing that further guidance ethical decision making regarding this topic is unnecessary: “Some people are just not good at collaboration, but I don’t think that is something that you still want to teach in the Master. I think people should know about themselves whether they are good in collaboration or not, and therefore it [discussing collaboration with peers during a course on RI] is kind of stupid.”

One of those three students did mention she was interested in peer feedback however, since she often did not receive high quality peer feedback even though she put a lot of effort into hers, which frustrated her.

These opinions are opposite the fourth interviewee and several of the students who completed the questionnaire, since they did all consider collaboration an important topic. The interviewee said that "since collaboration is such a key part of research for the rest of my life, addressing it is important", and the questionnaire respondents also stress the importance of the topic and at least hinted at failed collaborations. Examples are as follows: "Colaboration [sic] in groups is obligated in almost every course. Often some teammates do not deliver as much work as agreed on/expected." and "Collaboration is something that comes back in almost every course, often determines a large part of your grade and it is something that causes issues quite often."

References

This topic also received varied responses, with two students ranking it the lowest out of the four topics, one ranking it third and one ranking it second. The general consensus seems to be that while references are important and they are necessary for every course, they are not *as important* as the other topics ranked higher.

In terms of what they want to learn about it, two students wanted to explore trustworthiness of sources. One student however found this unnecessary, since they felt you already learn this during middle/high school and thus further guidance regarding ethical decision making is not necessary to discuss during a course on RI during the Master. This same student was interested in reference programs however, since they are easy to use and everyone is then on the same level for group projects.

One of the students did also mention a very interesting issue; the Graduate School of Teaching does have a required course that (among other things) teaches about references, but this course only occurs around halfway through the Master. Since references are required for every course, the students have already had to figure out referencing if they haven't learned it during their Bachelor, making a course on it have less use than if it was given immediately at the start of the Bachelor.

References were mentioned considerable less often by students during the questionnaire than the other topics, but the few students that did mention it, stressed its importance: "Incorrect/invalid refercing [sic] is most relevant to me, since referencing is probably an important part of writing my master thesis. And most of the time I just do what I think is logical but I have no idea if I'm following common practice or not. However, I could probably just ask my supervisor."

And "Use of literature: incorrect or invalid referencing

Many students have trouble with this, this is not very clearly explained in the bachelor but still very important in the master." are two examples.

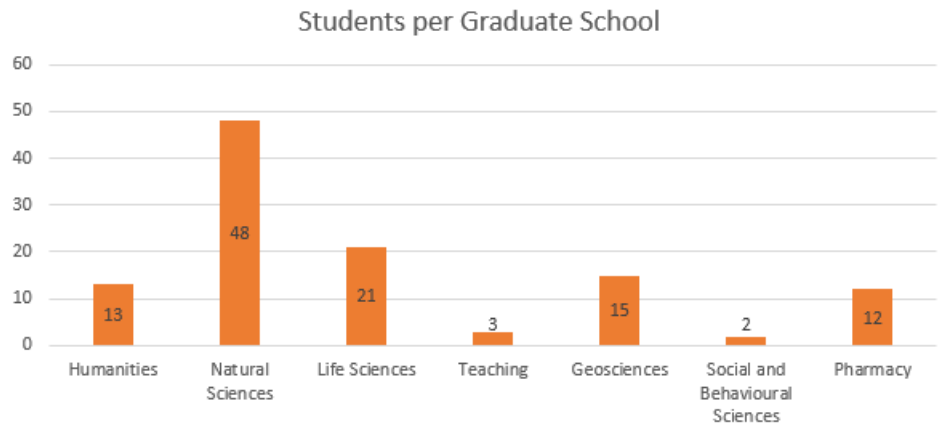
Graduate school differences

Respondents

The questionnaire asked students for their RMA, but during the analysis students were grouped by their Graduate Schools. Not only is this necessary for answering the Research sub-Question, it increased the amount of students per group which made analysis easier. Not all Graduate Schools belonging to Utrecht University were included or got respondents, only the Graduate Schools of Humanities, Natural Sciences, Life Sciences, Teaching, Geosciences, Social and Behavioural Sciences and Pharmacy were able to be analysed in this study.

Figure 6

Students per Graduate School



Note. The amount of students per Graduate School. The x-axis shows the name of the Graduate Schools, the y-axis the amount of students.

Out of the respondents, the most RMA students are from the Natural Sciences (48; 42.1%) and Life Sciences (21; 18.4%) Graduate schools. The least amount of students are from the Teaching (3; 2.6%) and Social and Behavioural Sciences (2; 1.8%).

For the interviewees, all four RMA students belonged to the Graduate School of Teaching.

Quantitative analysis

The sub-Question of my thesis concerns differences between Graduate Schools in articulating matters regarding RI. This section, like the one previously, will be split into the seven ethical dilemmas and the three sub-questions regarding 16 RI-related topics.

I have analysed the differences through the statistical test Kruskal-Wallis H, which is suited for ordinal data like this questionnaire. It gives each group (Graduate School in this case) a mean rank based on the size of their observations. When this rank is below the overall mean, the group has a lower size of observations, when a rank is above the size is higher than average. When there is a large difference in the ranking, the test gives a statistically significant result. This therefore means that there is a large difference in how students answered that question.

In this questoinnaire, I scored most relevant, often and most important with one, while not relevant and not important at all and never are scored with 5. When a Graduate School therefore is given a rank lower than the overall mean, the students belonging to that Graduate School thus consider that

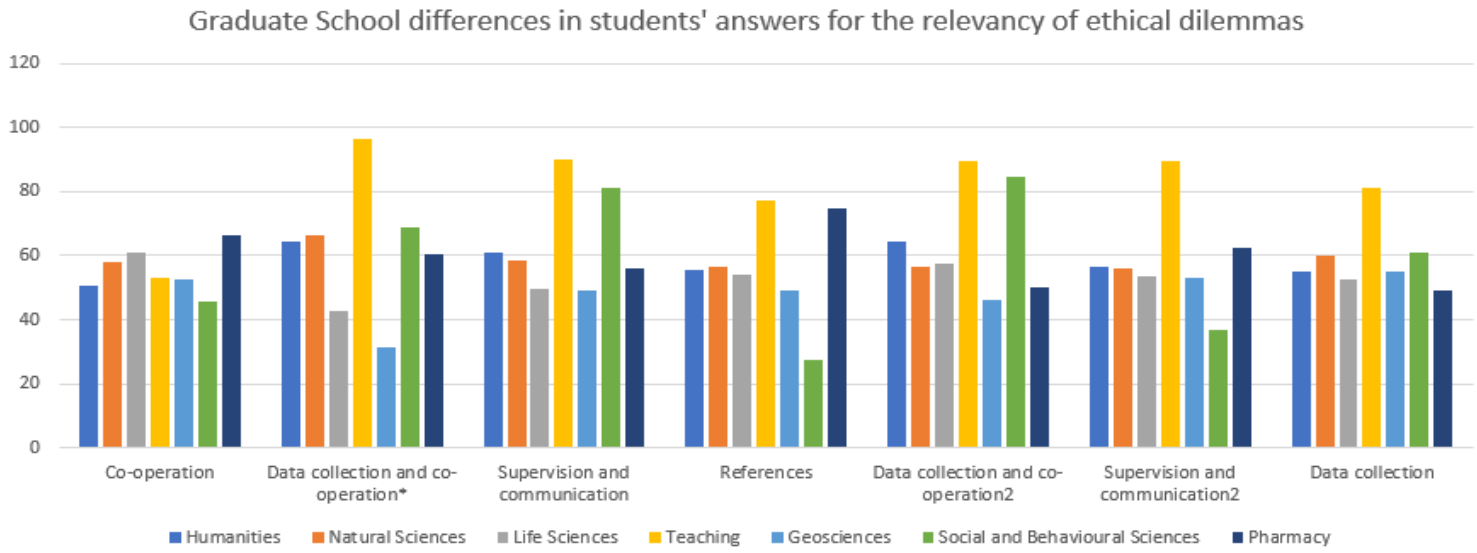
ethical dilemma or RI-related topic to be very important. When a Graduate School has a rank way higher than the overall mean, the students belonging to that Graduate School then consider that ethical dilemma or RI-related topic to not be very important. However, these rankings are abstract numbers, so in my analysis I will focus on Graduate Schools that rank way lower than the overall mean (**negative** outliers), and those that rank way higher (**positive** outliers).

I have also decided to showcase not only the statistically significant ethical dilemmas and RI-related topics, but the results for them all. This is mainly because only three students belong to the Graduate School of Teaching and two to Social and Behavioural Sciences, meaning their results and thus any statistical significance can be unreliable. I also want to be able to show general trends for certain Graduate Schools, not just how the Graduate Schools rank when an ethical dilemma or RI-related topic is statistically significant.

Ethical dilemmas analysis

Figure 7

The differences between Graduate Schools for the ethical dilemmas



Note. The differences between Graduate Schools for the ethical dilemmas. The x-axis shows the topic(s) of the ethical dilemmas according to Graduate School, see legend. Any ethical dilemma with a star (*) at the end is statistically significant. The y-axis shows the ranking of each Graduate School per ethical dilemma, which is an arbitrary number. One should look at how the rankings compare to each other instead. Graduate Schools with a **lower** than average ranking have students who think the ethical dilemma is relevant, and Graduate Schools with a **higher** than average ranking have students who think the ethical dilemma is less relevant.

As seen in the figure above, the first data collection and co-operation dilemma is the only statistically significant ethical dilemma, probably due to the low ranking of the Geosciences Graduate School, and the high ranking of the Teaching Graduate School. This means that this is an important topic for Geosciences, and not one for Teaching.

Other interesting trends are that the Humanities and Natural Sciences Graduate Schools have very similar rankings, so their interest in these ethical dilemmas is roughly equal to each other. Meanwhile the Teaching and Social and Behavioural Sciences Graduate Schools typically have very high rankings, meaning that most of these ethical dilemmas are not very relevant for them. This could however be due to their low participants, rather than any lack of care about RI.

Finally, the ethical dilemma that seems to be answered the most similarly is the one about co-operation, so the overall student population seems to feel similarly about it. The one with the most varied rankings besides the statistically significant first data collection and co-operation dilemma, is probably the references one, so opinions are very divided regarding this topic.

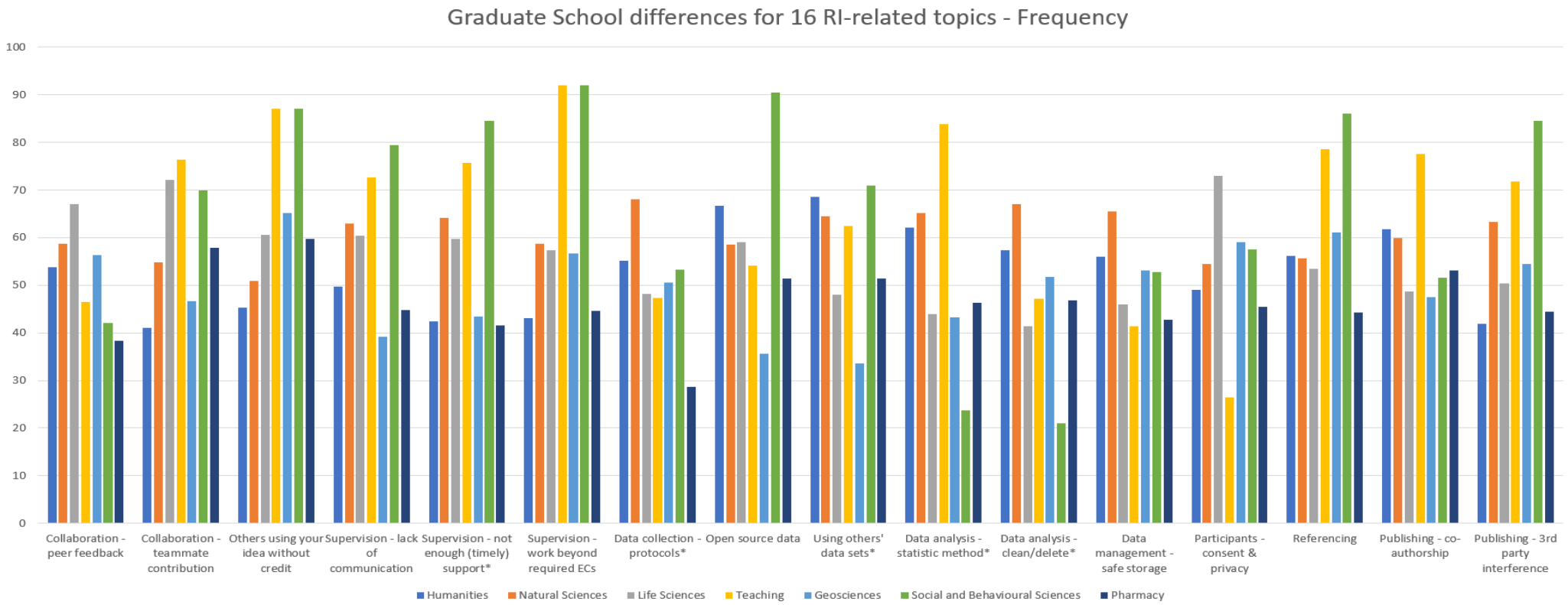
RI-related topics analysis

As mentioned above, I will separate the question regarding 16 RI-related topics into the three sub-questions for a better overview, and look at any statistically significant topics and interesting trends. The following pages are all in landscape orientation instead of portrait, to make the figures more legible.

Frequency

Figure 8

The differences between Graduate Schools for the 16 RI-related topics - Frequency



Note. The differences between Graduate Schools for 16 RI-related topics regarding frequency. The x-axis shows the 16 RI-related topics according to Graduate School, see legend. Any RI-related topic with a star (*) at the end is statistically significant. The y-axis shows the ranking of each Graduate School per RI-related topic, which is an arbitrary number. One should look at how the rankings compare to each other instead. Graduate Schools with a **lower** than average ranking have students who experience the RI-related topic often, and Graduate Schools with a **higher** than average ranking have students who experience the RI-related topic rarely.

As seen in the figure above, the statistically significant RI-related topics are supervision not enough (timely) support, data collection protocols, using others' data sets, data analysis statistic method and clean/delete, meaning there are large differences in how often students from different Graduate Schools experience these topics.

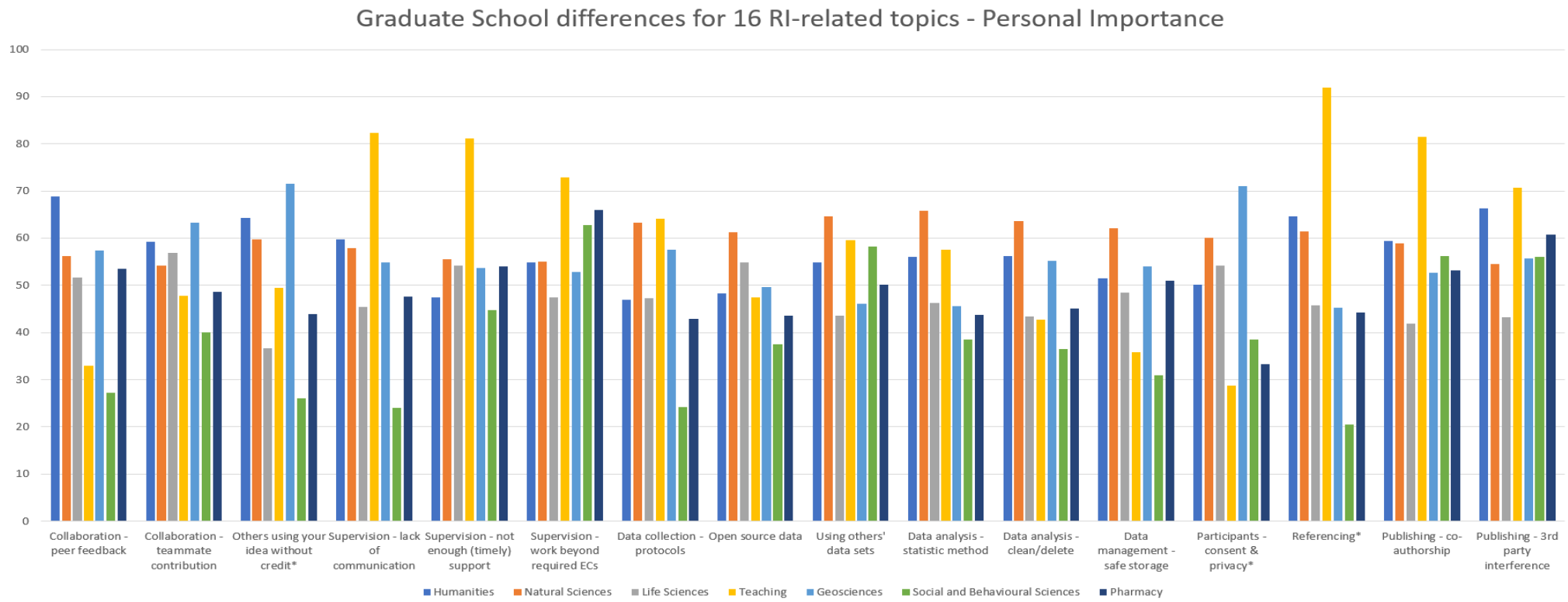
The predominant Graduate Schools with high rankings are once again the Teaching and Behavioural Science ones, again either due to rarely experiencing these topics or too low numbers. It seems like the Graduate School of Natural Sciences does not often have issues with data-related topics since they are all ranked rather high at those, whereas the Graduate School of Pharmacy is typically in the middle of the pack but often experiences issues with data collection protocols, since they have a low ranking there.

Overall data collection storage seems to be answered rather similarly, whereas the several statistically significant topics are of course all answered most differently and thus experienced at different rates during the RMA.

Personal importance

Figure 9

The differences between Graduate Schools for the 16 RI-related topics – Personal Importance



Note. The differences between Graduate Schools for 16 RI-related topics regarding personal importance. The x-axis shows the 16 RI-related topics according to Graduate School, see legend. Any RI-related topic with a star (*) at the end is statistically significant. The y-axis shows the ranking of each Graduate School per RI-related topic, which is an arbitrary number. One should look at how the rankings compare to each other instead. Graduate Schools with a **lower** than average ranking have students who think the RI-related topic is important, and Graduate Schools with a **higher** than average ranking have students who think the RI-related topic is less important.

As seen in the figure above, the statistically significant RI-related topics are others using your idea without credit, participants consent and privacy and referencing, meaning there are large differences in how important students from different Graduate Schools find these topics.

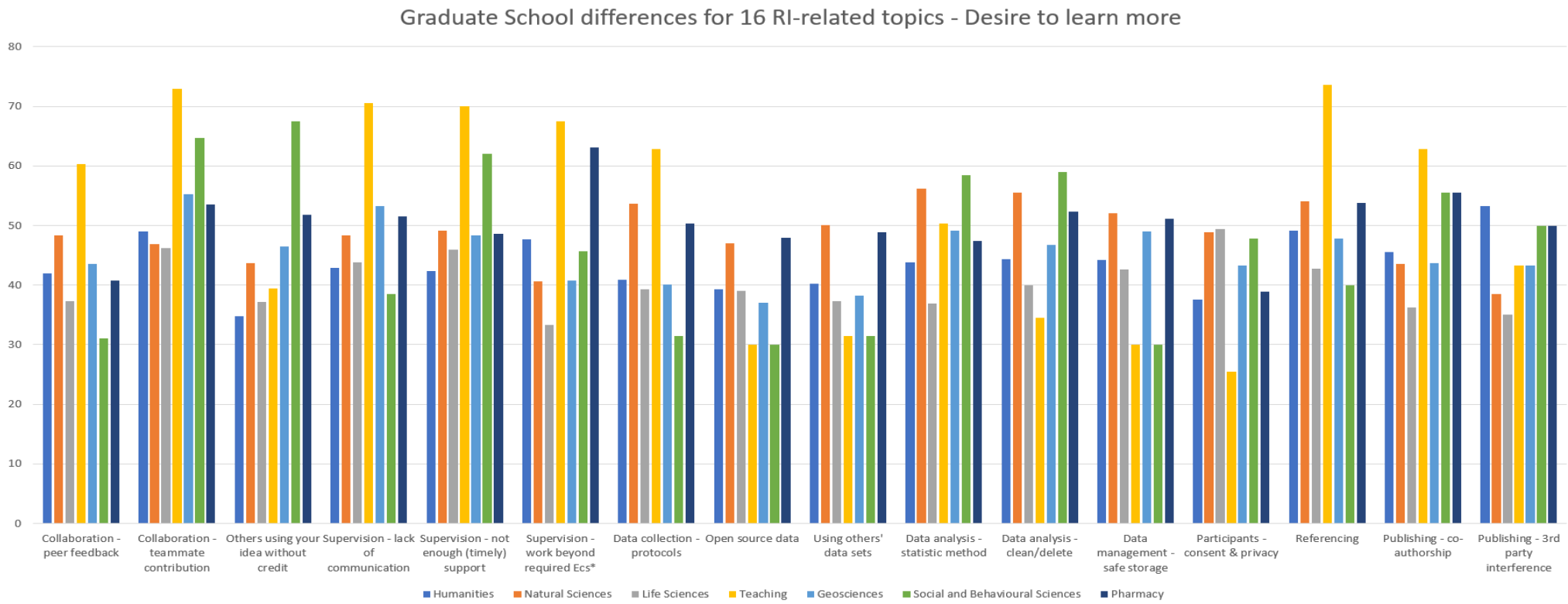
This sub-question has a lot fewer high rankings, but they are still mostly from the Graduate School of Teaching. There are now a lot more low rankings, mostly from the Graduate School of Social and Behavioural Sciences, meaning that a lot of students do find these RI-related topics important, even though they might not experience them frequently. The Graduate School of Geosciences seems very uninterested comparatively in RI-related topics such as others using your idea without credit and participants consent and privacy, while the Graduate Schools of Humanities and Natural Sciences once again seem to be answering pretty similarly to each other.

There are now also a few topics with no massive difference in rankings, such as collaboration teammate contribution, data analysis statistic method and clean/delete, even though the latter two were statistically significant in the Frequency sub-question. This means that frequency of experiencing an RI-related topic and personal interest in an RI-related topic are at least slightly uncorrelated. The three statistically significant topics in this question are of course ranked most differently, meaning the overall student opinion on how important these topics are, is divided.

Desire to learn more about

Figure 10

The differences between Graduate Schools for the 16 RI-related topics – Desire to learn more about



Note. The differences between Graduate Schools for 16 RI-related topics regarding students' desire to learn more about them. The x-axis shows the 16 RI-related topics according to Graduate School, see legend. Any RI-related topic with a star (*) at the end is statistically significant. The y-axis shows the ranking of each Graduate School per RI-related topic, which is an arbitrary number. One should look at how the rankings compare to each other instead. Graduate Schools with a **lower** than average ranking have students who want to learn more about the RI-related topic, and Graduate Schools with a **higher** than average ranking have students who don't want to learn more about the RI-related topic.

As seen in the figure above, the statistically significant RI-related topic is only supervision work required beyond required ECs, so students from different Graduate Schools answered this question rather similarly. This might however be due to the fact that the only two options for this question were 'Yes' and 'No', thus reducing the possible answer options massively.

This sub-question is a bit divided regarding especially low or high rankings, with a lot of high rankings from the Graduate Schools of Teaching and Social and Behavioural Sciences in the first few topics, and then two more in the final two few topics. The Graduate School of Teaching also has a lot of low rankings in the middle set of topics, it seems they really want to learn more about things like open source data and participants consent and privacy. The Graduate School of Pharmacy generally ranks pretty highly on most topics, thus maybe indicating that students from this Graduate School either are not particularly interested in learning more about these topics, or feel that their current education is good enough.

The RI-related topics ranked most uniformly are I think the two middle topics, so open source data and using others' data sets. The most diverse is, of course, the statistically significant one in supervision work beyond required ECs.

Qualitative analysis

As stated during the qualitative analysis for the first Research sub-Question, the RI-related topic of data differed between Graduate Schools. The interviewed students, all belonging to the Graduate School of Teaching, expressed a lot of interest in and need for RI education regarding qualitative data. One student for instance found the following difficult: "I think especially the confidentiality of information and how to handle that appropriately. What can you ask, what can't you, what falls under which type of consent? And also, what can you ask of people, participants? When do you want to collect too much data and it's not really relevant, and when is it fine, so people are okay with participating in your research?"

This was in stark contrast to the questionnaire respondents, who indicated their interests lay more with quantitative data topics, such as statistical analysis: "data analysis: unsure about statistical method; this is what I fear most for my upcoming master thesis".

One other difference is that the Maths RM students do not need to collect or work with data:

"Given that I do pure maths without any data input, I can only really think of problematic situations in communication with supervisors."

Finally, the RMA students belonging to the Geosciences Graduate School receive little to no education regarding RI, referencing and data:

"I would like to say that in my Earth Sciences bachelor and master there were hardly ANY lectures on integrity, finding and using literature, and statistics. I have heard that this is present in other programs. When I followed an anthropology minor, considerable attention was given to these topics, and I learned more about integrity and literature during a few of those lectures than during the whole of my other studies."

Discussion and conclusion

This thesis aimed to help better RI education, through reinvigorating RMA students' interest in the subject. The research would do so by identifying two things: which topics within RI are relevant to students so ethical dilemmas for RI education could be better formulated or selected, and whether Graduate Schools are a determinant for RMA students' interest in certain RI-related topics.

The questionnaire and the interviews both showed that the four topics most relevant to RMA students are supervisor (lack of support and communication), data (ethical handling of both qualitative and quantitative data), collaboration with peers (unequal contribution and receiving low-quality peer-feedback) and (incorrect) referencing.

Graduate Schools were an important determinant for RMA students' interest in the data topic, with the students from the Graduate School of Teaching more interested in ethical handling of qualitative data, and the RMA students of the other Graduate Schools more interested in quantitative data.

Other differences between Graduate Schools are that students from Maths-related RMAs do not collect their own data, making ethical dilemmas regarding this topic uninteresting for them, while students from the Graduate School of Geosciences expressed heavy interest in any education regarding RI, referencing and data.

Another topic that students are divided on is the topic collaboration with peers. Several of the interviewed students expressed that they find further guidance on this topic and especially on how to deal with unequal contribution unnecessary, since they have had to collaborate in various groups for years now. The questionnaire however showed that in the general RMA student population, collaboration with peers is a very relevant and important topic to them.

All these differences will make it difficult to make a coherent RI course meant to interest and engage every student. My advice or answer to the main Research Question is therefore as follows: introduce flexibility in a RI course, in the form of several ethical dilemmas for each of the four main topics which address various sub-topics, such as one which is more focused on qualitative data, and one more focused on quantitative data. The RI teacher can then select appropriate ones per lesson and group of students. The teacher can even sub-divide students based on interest in a particular ethical dilemma, to avoid the problem of students getting bored with for instance an ethical dilemma related to collaboration with peers.

There are three other important things to note. First, there is no clear correlation between the frequency of experiencing a topic in the RMA and how important the students find it, especially shown in the Graduate School analysis. Both factors are however important, with this study showing that the personal importance students hold in a topic is correlated to their desire to learn about the topic (the four highest and two lowest scoring RI-related topics were the same), but students also have interest in topics that they have experienced (see again the difference in qualitative versus quantitative data per Graduate School). This means both factors need to be taken into account when selecting appropriate topics.

Secondly, the sub-topics contained in the ethical dilemma can lead to different feelings of relevance in students. This can for instance be seen in the difference between the first and second data collection and co-operation ethical dilemma, which addressed only slightly different sub-topics but were regarded very differently.

Finally, the experiences between Graduate RMA students and PhD students differ. Both Jensen et al. (2018) and Hofmann and Holm (2019) found that the most common conflict for PhD students

regarded pressure over authorship, with respectively over 20% and 30% of students experiencing it. In this study, 37.8% of students often or sometimes experienced issues regarding co-authorship, which is more than the other studies found, but it was not the topic RMA students experienced *most* frequently. This difference might simply be explained as a difference of method, e.g. the two studies did not ask for PhD students' experiences regarding all the RI-related topics in this study, or my questionnaire did not ask what *kind* of co-authorship problems the RMA students had experienced. However, the difference in percentages does hint to a differing experience between RMA students and PhD students.

Limitations and future research

The main limitations of this study concern the second Research sub-Question regarding Graduate Schools, both for the questionnaire, as well as the interviews. The questionnaire only had three students from the Graduate School of Teaching participate, and two from Social and Behavioural Sciences. This meant that these two Graduate Schools were under represented, and that smaller differences between Graduate Schools might have gone unnoticed.

To ameliorate this problem, I chose to portray all the results for the Kruskal-Wallis H test showing differences between Graduate Schools, not just the statistically significant question items. These items may only be statistically significant due to the low number of participants of certain Graduate Schools creating easy outliers, not any actual difference in reality. Thus, it is more valid to portray all the data and notice any trends in Graduate School responses, not just focus on statistical significance.

As for the interviews, all interviewed students belonged to the Graduate School of Teaching, which might skew my results in this section of my thesis. However, for my full qualitative data analysis I also included qualitative data from the questionnaire, where the other Graduate Schools were more represented. It is of course not a perfect fix, but it does give a better overview of the whole student population.

With regard to future research, I see three main directions: further collecting data on students' preferences or determinants for effective RI education, discovering further differences between PhD and RMA students and how to handle those differences in a course on RI, or researching how well the recommendations formulated in this thesis work in practice. For the first, one can think of trying to answer the sub-question regarding differences between Graduate Schools once more, though this time with more representative numbers of participants per Graduate School.

Other possible determinants for effective RI education could be gender or socio-economic background. For gender, one can think about differences in interest regarding RI, both in what a certain gender holds as important and thus wants to know about; and what they have experienced in their RMA and thus need help with. As for socio-economic background, people from a poor socio-economic background may hold totally different views on authority and its abuses than those coming from a rich socio-economic background, so this might also be interesting to investigate.

This study showed that there is a difference between what RMA students' and PhD students' experience. This introduces a dilemma: how should a course on RI focus on RMA students' interests, when some of those might not be (very) relevant in the rest of their research career? The second direction would therefore include investigating precise differences between these two target groups, and then deciding which ones matter in the practice of teaching RMA students about RI.

I am however most interested in seeing if my recommendations work in practice, so the third direction, through creating new RI educational material and testing its effectiveness via intervention research. To quickly reiterate, ethical dilemmas would play a central role, centred around the topics of supervisor (lack of support and communication, abuse of authority), data (ethical handling of quantitative and qualitative data), co-operation with peers (unequal contribution, receiving low-quality peer-feedback) and referencing (how to properly and ethically reference). A very important design rule would be flexibility, since students both personally and per Graduate School have different interests in certain fields of RI, which the newly developed material should take into account in order to keep students more interested and engaged.

References

About Elevate Academy. (2021, December 6th). Elevate. Consulted on March 30th 2022, from https://elevatehealth.eu/about-elevate-academy/

ALLEA – All European Academics (2017). *The European code of conduct for research integrity* (Revised Ed.). ALLEA – All European Academics. https://allea.org/code-of-conduct/

Antes, A. L., & DuBois, J. M. (2014). Aligning Objectives and Assessment in Responsible Conduct of Research Instruction. *Journal of Microbiology & Biology Education*, 15(2), 108–116. https://doi.org/10.1128/jmbe.v15i2.852

Bredahl-Jensen, L. & Kyvik, K. & Leth-Larsen, R., & Eriksen, M. (2018). Research integrity among PhD students within clinical research at the University of Southern Denmark. *Danish medical journal*, 65.

Broad, W. J. (1980). Would-be academician pirates papers. *Science*, 208: 1438–1440.

Connelly L., Burbach B.E., Kennedy C., Walters L., Escape room recruitment event: Description and lessons learned, *Journal of Nursing Education*, 57 (3) (2018), pp. 184-187

Hofmann, B., & Holm, S. (2019). Research integrity: environment, experience, or ethos? *Research Ethics*, 15(3–4), 1–13. https://doi.org/10.1177/1747016119880844

Kalichman, M. (2013). A Brief History of RCR Education. *Accountability in Research*, 20(5–6), 380–394. https://doi.org/10.1080/08989621.2013.822260

Kalichman, M. W., & Friedman, P. J. (1992). A pilot study of biomedical trainees' perceptions concerning research ethics. *Academic Medicine*, 67: 769–775

Langlais, P. J., & Bent, B. J. (2013). Individual and Organizational Predictors of the Ethicality of Graduate Students' Responses to Research Integrity Issues. *Science and Engineering Ethics*, 20(4), 897–921. https://doi.org/10.1007/s11948-013-9471-2

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. 1992. *Responsible Science: Ensuring the Integrity of the Research Process: Volume I*. Washington, DC: The National Academies Press. https://doi.org/10.17226/1864.

ORI - The Office of Research Integrity. (2022). Homepage. Consulted on April 2nd, 2022, from https://ori.hhs.gov/

Relman, A. S. (1983). Lessons from the Darsee Affair. *New England Journal of Medicine*, 308: 1415–1417

Resnik, D. B. (2014). Editorial: Does RCR Education Make Students More Ethical, and Is This the Right Question to Ask? *Accountability in Research*, 21(4), 211–217. https://doi.org/10.1080/08989621.2013.848800

Shephard, K., Trotman, T., Furnari, M., & Löfström, E. (2015). Teaching research integrity in higher education: policy and strategy. *Journal of Higher Education Policy and Management*, 37(6), 615–632. https://doi.org/10.1080/1360080x.2015.1102823

Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics*, 12(1), 53–74. https://doi.org/10.1007/s11948-006-0006-y

Swazey, J. P., & Anderson, M. S. (1998). Mentors, Advisors, and Role Models in Graduate and Professional Education. In Rubin, E.R., ed., Mission Management. Washington, D.C.: Association of Academic Health Centers.

Van den Hoven, M., Application form Escape Room Responsible Conduct of Research, submitted to *the Utrecht Education Incentive Fund* before the 24th of January 2020.

Veldkamp, A., van de Grint, L., Knippels, M.-C. P. J., & van Joolingen, W. R. (2020). Escape education: A systematic review on escape rooms in education. *Educational Research Review*, 31, 100364. https://doi.org/10.1016/j.edurev.2020.100364

Appendix A – The questionnaire – all ethical dilemmas and RI-related topics

The second part (out of four) of the questionnaire poses seven ethical dilemmas to the students, and they are asked to answer them on a 5-point Likert scale, with 1 being highly relevant, 2 relevant, 3 neutral/don't know, 4 not very relevant, and 5 not at all relevant. They are:

1. In a group project, I notice that one of my group mates is not contributing fairly. (S)he misses every deadline and the work is a bit sloppy. The grade will be 60% of my final grade for this course. I am not sure what to do.
2. Due to circumstances out of my control, I cannot collect my own data, and I am relying on datasets of previous students. Yet, the dataset that I receive is not of high quality. I doubt whether I can use it.
3. A supervisor has recently been appointed to me for my final thesis project, but (s)he is not responding to my emails. I really want to get started with my project.
4. I am finishing a paper, and need a few references that confirm my statements. I choose a few articles that only briefly mention something about my paper topic, but I don't have the time to search for new references, so I decide to use them anyway.
5. My supervisor suggests that I use the dataset that another student is also working on. When asking for the dataset, (s)he does not respond and when I finally get in touch, (s)he refuses to share the dataset. I am not sure what to do.
6. I have been working on my thesis for a while, but it seems like my supervisor has less and less time for me. (s)he is difficult to contact, barely or just doesn't read my work, and any useful commentary is too late to be of help.
7. I will collect my own data, and one of my fellow students mentions that I need to write a data management plan in advance. Yet, I already made appointments for the interviews with my respondents. I am not sure what to do.

The third part of the questionnaire asks three things about sixteen different RI-related topics. These are: 1. How often they come across the topic in their Master, 2. How important the topic is to them personally, and 3. Whether they would like to learn how to deal with them. The first and second questions can be answered with 5-point Likert scales, with the scales respectively being 1 often, 2 sometimes, 3 don't know, 4 rarely, 5 never; and 1 highly important, 2 important, 3 neutral/don't know, 4 not very important, 5 not important at all. The third question only has the option of yes (1) or no (2).

The RI-related topics are as follows:

1. Collaboration: peer feedback, unconstructive or too little
2. Collaboration: teammate contributes less than agreed upon
3. Somebody else using your idea without giving you credits
4. Supervision: lack of communication
5. Supervision: not receiving enough or timely support for your project
6. Supervision: being asked to do extra work beyond the ECs you receive
7. Data collection: how to plan it out, formulating proper protocols
8. Use of open source data and repositories
9. Quality and integrity of datasets you use/borrow from others: can you trust them?
10. Data analysis: unsure about right statistical method and coding of data
11. Data analysis: When can you legitimately delete or clean data from a dataset?
12. Data management: storing data safely

13. Interaction with participants: arranging informed consent, privacy
14. Use of literature: incorrect or invalid referencing
15. Publishing/presenting results: Whom to put as co-author on your report?
16. Conflict of interest: third party prohibits publication of results

Appendix B – Edits to raw data file

I exported the raw data file of the survey from Qualtrix to SPSS and later Excel. To make the raw data file easier to work with, I made a few edits. First I deleted the variables that were unnecessary, these were as follows:

1. The date the student started filling in the survey (StartDate).
2. The date the student finished filling in the survey (EndDate).
3. The status of the student's computer (Status).
4. The student's progress in percentages (Progress).
5. The time it took the student to fill the survey in (Duration_in_seconds).
6. Whether the student had finished or not (Finished).
7. The date the response was recorded (RecordedDate).
8. The method of distribution the student took to fill in the survey (DistributionChannel).
9. The language the student used to fill in the survey (UserLanguage).

I also removed all students who did not fill in the survey, which lead to the deletion of 86 responses, meaning the amount of responses given by students dropped from 200 to 114.

I also created five new variables in the SPSS file, which are as follows:

1. Graduate_schools, which takes the answers from Q2 and sorts them into specific graduate schools with 1=Humanities, 2=Natural Sciences, 3=Life Sciences, 4=Teaching, 5=Geosciences, 6=Social and Behavioural Sciences, and 7=Pharmacy
2. Q10_1_all, which combines the answers of a student to all 16 items of the Frequency sub-question of Q10.
3. Q10_2_all, which combines the answers of a student to all 16 items of the Personal Importance sub-question of Q10.
4. Q10_3_all, which combines the answers of a student to all 16 items of the Learn more about it? sub-question of Q10.
5. Most_relevant, which takes the answers to the open question Q11, which of the topics of Q10 is most relevant to the students, and sorts them into categories. 1=Data (from analysis to informed consent as part of data collection), 2=Supervisor, 3=Collaboration with peers, 4=Publication and conflict of interest, 5=Plagiarism, 6=Referencing, 7=None/no specifics, 8=Any combination of data, supervisor, collaboration and referencing.

The Excel file only contains the variables Graduate_schools and Most_relevant.

Appendix C – Interview questions

Hi, my name is Romy Zandstra and for my Master thesis I am collaborating on a project aiming to make research integrity more fun and interesting for Research Master students. I am specifically investigating which research integrity topics students find important or interesting. The first part of the research consisted of a survey from which we identified a few important topics, and with these interviews I want to collect some more qualitative data to see *why* these topics are important to students. I am going to ask you questions about the following four subjects: supervisor, data, collaboration with peers and referencing. I would like to both know your personal experience with these topics, and whether or not you find them important.

Before we begin I would like to ask you two consent questions:

1. Do you give consent to participate in this interview?
2. Do you give consent to me recording this interview?

Then I want to ask two questions regarding personal information:

1. Which Master are you following and to which Graduate School does it belong to?
2. What is your gender?

Now for the core of the interview, the research integrity topics supervisor, data, collaboration with peers and referencing. Do you have a preference to discussing one first, or shall we just go down the list?

Supervisor:

1. Did you have a supervisor during your Master?
2. Was this supervisor for your thesis, internships, normal courses or a combination of the three?
3. How much influence did this supervisor have? Prompt: think about grades, advice, overall experience of the course/internship/thesis.
4. Did you have a particularly memorable experience with a supervisor? What was this experience, and was it positive or negative?
5. Can you pinpoint why this experience was positive or negative?
6. Do you think the topic of supervisor is important enough to include in a course on research integrity? Why (not)?
7. What do and do you not want to learn? Several sub-topics are the power balance between supervisor and a student, or what to do when your supervisor does not keep in frequent contact.

Data:

1. Do you need to collect data in your Research Master?
2. Is there a course during your Research Master which explains how to work with data?
3. Do you already have experience with data? Collecting it, analysing it, etc.
4. If so, what do you remember most clearly and which part did you struggle most in handling data?
5. Can you pinpoint why you remembered that part so clearly and why you struggled with that part of data?
6. Do you find the topic of data important enough to include during a course on Research Integrity? Why (not)?

7. What do you and do you not want to learn about? Several sub-topics besides directly handling data are open source data, privacy of research participants and dealing with ethical commissions.

Collaboration with peers:

1. Do you need to collaborate with your peers during your RM?
2. Is collaborating with your peers a key part of your RM in your opinion?
3. Do you have a particularly memorable experience with collaboration? What was this experience, and was it a positive or negative one?
4. Can you pinpoint why this experience was positive or negative?
5. Do you think the topic of collaboration with peers is important enough to include in a course on research integrity? Why (not)?
6. What do you and do you not want to learn about? Several sub-topics are dealing with students who don't do their work, how to actually create an effective collaboration, or how to give and receive peer feedback.

References:

1. Do you need to use references during your RM and if so, how often?
2. Have you had instruction on how to deal with references?
3. Which kind of references do you use most often? Think about books, websites, academic papers, etc.
4. Do you use a reference program?
5. Have you have an experience with references that was particularly memorable? What was this experience, and was it positive or negative?
6. Can you pinpoint why the experience was positive or negative?
7. Do you consider the topic of references important enough to include in a course on research integrity? Why (not)?
8. What do you and do you not want to learn about? Several sub-topics are for instance how to deal with reference programs, how different reference styles work, or how to deal with different sources.

Finally, I would like you to rank the four topics, from most important to least important to you, so supervisor, data, collaboration with peers and references.

Can you explain why you ranked the topics this way?